

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 : Search time 40.3636 Seconds
(without alignments)
23.594 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29
Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	6	21	AA03860
2	31	100.0	40	21	AA03864
3	31	100.0	60	22	AA09415
4	31	100.0	159	20	AA43626
5	31	100.0	1204	21	AG48838
6	31	100.0	1216	21	AG48837
7	31	100.0	1248	21	AG48836
8	30	96.8	220	22	AAE13069
9	30	96.8	389	22	AAU34465

10	30	96.8	389	22	AAU38301	Salmonella typhi c
11	29	93.5	28	21	AB44520	Plant viral move
12	29	93.5	69	21	AB44518	Plant viral move
13	29	93.5	145	21	AB44507	Plant viral move
14	29	93.5	697	22	ABG25300	Novel human diago
15	29	93.5	751	24	ABR41404	Human DTHP zinc f
16	28	90.3	44	19	AY20541	Human neurofilamen
17	28	90.3	81	22	ABG20244	Novel human diago
18	28	90.3	89	24	AB82574	H. influenzae BASB
19	28	90.3	92	22	AAU01706	Propeptide of cono
20	28	90.3	106	22	ABG20246	Novel human diago
21	28	90.3	117	23	ABP06804	Human ORFX protein
22	28	90.3	173	23	AB57434	Human secreted pro
23	28	90.3	174	23	AB57435	Human secreted pro
24	28	90.3	178	21	AG30897	Arabidopsis thalia
25	28	90.3	280	22	AB58792	Drosophila melanog
26	28	90.3	311	19	AAW58530	Human peroxisomal
27	28	90.3	311	20	AAW08604	Human PXT protein
28	28	90.3	311	22	AG77835	Protein sequence o
29	28	90.3	311	22	AB73420	Human peroxisomal
30	28	90.3	331	22	ABG20249	Novel human diago
31	28	90.3	341	21	AB36830	37KD protein. Bab
32	28	90.3	392	22	AG64743	Human sodium neuro
33	28	90.3	554	20	AY35719	C. pneumoniae prot
34	28	90.3	682	23	AB49246	Listeria monocytog
35	28	90.3	729	23	AG21215	Protein sequence i
36	28	90.3	730	22	AG67159	Amino acid sequenc
37	28	90.3	730	22	AG72908	Human NTF7 protein
38	28	90.3	730	23	AB77168	Human GABA transpo
39	28	90.3	730	23	AAE14404	Human neurotransmi
40	28	90.3	759	24	ABJ25504	Aspergillus fumiga
41	28	90.3	860	14	AA42962	Bovine MTP. Bos t
42	28	90.3	869	24	ABJ26104	Aspergillus fumiga
43	28	90.3	914	23	AAE18115	Human MDM2 binding
44	28	90.3	2248	22	AB64494	Drosophila melanog
45	27	87.1	46	22	ABG49213	Human liver peptid

ALIGNMENTS

RESULT 1
AAB03860
ID AAB03860 standard; peptide; 6 AA.
XX
AC AAB03860;
XX
XX 20-OCT-2000 (first entry)
DT
XX
DE Peptide #1 used in expression vector construction.
XX
KW Expression vector; active target protein; easy isolation.
XX
OS Unidentified.
XX
PN WO200031284-A1.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-JP06474.
XX
PR 20-NOV-1998; 98JP-0331515.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;
XX
DR WPI; 2000-411958/35.
XX
PT Protein expression vectors containing specific sequences with cloning
PT sites into which nucleic acids insert, for isolation of target proteins
XX without excess amino acids -
XX

PS Claim 5; Page 21; 44pp; Japanese.

XX The invention relates to a protein expression vector from which mature
CC and active target proteins can be easily isolated. N and C-terminal amino
CC acids are kept to a minimum to ensure that easy isolation from the vector
CC is possible. The protein expression vector comprises a secretory signal
CC nucleic acid sequence and, in the 3'-downstream side, a Tag nucleic acid
CC sequence, a scissile nucleic acid sequence and a cloning site into which
CC a nucleic acid sequence encoding a target protein can be inserted. Also
CC included in the invention are a host cell transformed by the vector, a
CC process for producing a target protein using the vector or host cell, a
CC recombinant fusion protein containing the amino acid sequence of the
CC target protein, and a process for producing a target protein. The
CC expression vector is used for the production of mature and active target
CC protein. The present sequence represents a peptide used in the
CC construction of the vector of the invention.

XX Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKGL 6
| | | | |
DB 1 LVHGKGL 6

RESULT 2

AAB03864
ID AAB03864 standard; Protein; 40 AA.

XX AAB03864;

XX 20-OCT-2000 (first entry)

XX Active human neurosin amino acid sequence.

XX Expression vector; active target protein; easy isolation; neurosin;
KW human.

XX Homo sapiens.

XX WO2000031284-A1.

XX 02-JUN-2000.

XX 19-NOV-1999; 99WO-JP06474.

XX 20-NOV-1998; 98JP-0331515.

XX (FUSO) FUSO PHARM IND LTD.

XX Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;

XX WPI; 2000-411958/35.

XX N-PSDB; AAA62817.

PT Protein expression vectors containing specific sequences with cloning
PT sites into which nucleic acids insert, for isolation of target proteins
PT without excess amino acids -

XX Example 3; Page 42; 44pp; Japanese.

XX The invention relates to a protein expression vector from which mature
CC and active target proteins can be easily isolated. N and C-terminal amino
CC acids are kept to a minimum to ensure that easy isolation from the vector
CC is possible. The protein expression vector comprises a secretory signal
CC nucleic acid sequence and, in the 3'-downstream side, a Tag nucleic acid
CC sequence, a scissile nucleic acid sequence and a cloning site into which
CC a nucleic acid sequence encoding a target protein can be inserted. Also
CC included in the invention are a host cell transformed by the vector, a
CC process for producing a target protein using the vector or host cell, a

CC recombinant fusion protein containing the amino acid sequence of the
CC target protein, and a process for producing a target protein. The
CC expression vector is used for the production of mature and active target
CC protein. The present sequence represents active human neurosin. Neurosin
CC is used as the target protein in an example of the vector of the
CC invention.

XX Sequence 40 AA;

Query Match 100.0%; Score 31; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKGL 6
| | | | |
DB 24 LVHGKGL 29

RESULT 3

AAM94415
ID AAM94415 standard; Protein; 60 AA.

XX AAM94415;

XX 21-NOV-2001 (first entry)

XX Human reproductive system related antigen SEQ ID NO: 3073.

XX Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.

XX Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179055.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

XX 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX N-PSDB; AAL00385.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 11; SEQ ID NO 3073; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
SQ Sequence 60 AA;
Query Match 100.0%; Score 31; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVHGKL 6
|||||
Db 44 LVHGKL 49
RESULT 4
AA43626
ID AAY43626 standard; Protein; 159 AA.
XX
AC AAY43626;
XX
DT 26-JAN-2000 (first entry)
XX
DE A portion of the starch R1 phosphorylation protein.
XX
KW Starch R1 phosphorylation protein; starch degradation;
KW plant starch biosynthesis.
XX
OS Brachythemium oxyclocladon.
XX

```

FH Key Location/Qualifiers
FT Misc-difference 110 /note= "unidentified amino acid"
FT Misc-difference 134 /note= "unidentified amino acid"
FT Misc-difference 137 /note= "unidentified amino acid"
FT Misc-difference 149..150 /note= "unidentified amino acids encoded by ATTCAAG"
FT Misc-difference 155 /note= "unidentified amino acid"
FT FT
XX
XX WO9953072-A1.
PN
XX
XX 21-OCT-1999.
PD
XX
XX 08-APR-1999; 99WO-US07639.
XX
XX 09-APR-1998; 99US-0081143.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA
XX Cressman RF, Allen SM;
PI
XX
XX WPI; 1999-620435/53.
DR N-PSDB; AAZ30159.
XX
XX New starch R1 phosphorylation protein homologues -
PT
XX Claim 1; Page 38; 54pp; English.
PS
XX
CC The present sequence represents a starch R1 phosphorylation protein.
CC The protein is involved in starch degradation. The specification
CC also describes a chimeric gene encoding all or a portion of the
CC starch R1 phosphorylation protein, in the sense or antisense
CC orientation, where expression of the chimeric gene results in
CC production of altered levels of starch R1 phosphorylation protein
CC in transformed plants or cells. The protein facilitates studies of
CC starch degradation in plants, and provides tools for the genetic
CC manipulation of starch biosynthesis in plants.
XX
XX Sequence 159 AA;
SQ
Query Match 100.0%; Score 31; DB 20; Length 159;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 LVHGKL 6
Db 138 LVHGKL 143
RESULT 5
AAG48838
ID AAG48838 standard; Protein; 1204 AA.
XX
XX AAG48838;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 61715.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF

```

99US-0121825.
99US-0123180.
99US-0123548.
99US-0125788.
99US-0126264.
99US-0126785.
99US-0127462.
99US-0128234.
99US-0128714.
99US-0129845.
99US-0130077.
99US-0130449.
99US-0130510.
99US-0130891.
99US-0131449.
99US-0132048.
99US-0132407.
99US-0132484.
99US-0132485.
99US-0132486.
99US-0132487.
99US-0132863.
99US-0134256.
99US-0134218.
99US-0134219.
99US-0134221.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 31; DB 21; Length 1204;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 47 LVHGKL 52

RESULT 6

AAG48837
ID AAG48837 standard; Protein; 1216 AA.

XX AC AAG48837;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61714.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.

OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 23-MAR-1999; 99US-0123548.

XX PR 25-MAR-1999; 99US-0125788.

XX PR 29-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134258.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      100.0%; Score 31; DB 21; Length 1216;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LVHGKL 6
Db      59 LVHGKL 64

RESULT 7
AAG48836
ID  AAG48836 standard; Protein; 1248 AA.
XX
AC  AAG48836;
XX
DT
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 61713.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0123788.
PR  29-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130449.
PR  23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
```

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 31; DB 21; Length 1248;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 91 LVHGKL 96

RESULT 8

AAE13069

ID AAE13069 standard; Protein; 220 AA.

XX AAE13069;

XX 28-JAN-2002 (first entry)

DE Brugia malayi (BM)-Asp protein.

XX Angiogenesis; Ov-ASP; therapy; circulatory disorder; vascular disorder;
congenital heart disease; myocardial disease; pericardial disease; wart;
cerebrovascular ischaemia; veno-occlusive disease; myocardial ischaemia;
coronary artery disease; diabetic retinopathy; inflammatory disease;
wound healing; duodenal ulceration; rheumatoid arthritis; psoriasis;
periodontitis; dermatological; cutaneous malignancy; Kaposi's sarcoma;
pyogenic granuloma; cancer; onchocerciasis; River Blindness; neoplasia;
vasotropic; cardiac; antiparasitic; ophthalmological.

OS Brugia malayi.

XX Key Location/Qualifiers

FT Peptide 1..16

FT Protein /label= Signal_peptide

FT /label= Mature_Em-Asp_protein

XX WO200174385-A1.

XX 11-OCT-2001.

XX 27-MAR-2001; 2001WO-US09798.

XX 03-APR-2000; 2000US-0541759.

XX (NYBL-) NEW YORK BLOOD CENT INC.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX (UABR-) UAB RES FOUND.

XX Lustigman S, Pearlman E, Unnasch TR;

XX WPI; 2001-662950/76.

XX Inducing angiogenesis in a tissue using the Ov-ASP protein isolated
PT from the nematode *Onchocera volvulus* is useful to treat circulatory or
PT vascular disease such as ischemia -
XX Disclosure; Fig 1; 37pp; English.

XX The present invention relates to a method for inducing angiogenesis in a
XX tissue. The method comprising contacting the tissue with Ov-ASP. The Ov-
XX ASP molecules are used to treat circulatory or vascular disorders,
XX particularly ischemia, congenital heart disease, myocardial disease or
XX pericardial disease, more particularly cerebrovascular ischaemia, veno-
XX occlusive disease or myocardial ischaemia, especially coronary artery
XX disease. The invention is also used to treat cancer, diabetic
XX retinopathy and inflammatory disease. Angiogenesis is also central to a
XX number of pathological processes, including abnormalities of wound
XX healing in diseases such as diabetes and duodenal ulceration; chronic
XX inflammatory disorders such as rheumatoid arthritis, psoriasis and
XX periodontitis; dermatological conditions such as cutaneous malignancy,
XX Kaposi's sarcoma, pyogenic granulomas and warts. Anti-Ov-ASP factors are
XX useful to treat onchocerciasis (River Blindness) or benign or malignant
XX neoplasia. The present sequence is Brugia malayi (Bm)-ASP protein.

XX Sequence 220 AA;
SQ

Query Match 96.8%; Score 30; DB 22; Length 220;
Best Local Similarity 83.3%; Pred. No. 88;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||
Db 43 LIHGKL 48

RESULT 9
AAU34465
ID AAU34465 standard; Protein; 389 AA.
XX
AC AAU34465;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #46.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Escherichia coli.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX WPI: 2001-611495/70.
DR N-PSDB; AAS52324.
DR
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 10058; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
XX *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 389 AA;
SQ

Query Match 96.8%; Score 30; DB 22; Length 389;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||
Db 65 LIHGKL 70

RESULT 10
AAU38301
ID AAU38301 standard; Protein; 389 AA.
XX
AC AAU38301;
XX
DT 14-FEB-2002 (first entry)
XX
DE *Salmonella typhi* cellular proliferation protein #192.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX *Salmonella typhi*.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX WPI: 2001-611495/70.
DR N-PSDB; AAS56160.
DR
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 13894; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

Seq Sequence 389 AA;

Query Match 96.8%; Score 30; DB 22; Length 389;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 Db 65 LIHGKL 70
 I:|||||

RESULT 11

AAB44520
 ID AAB44520 standard; Protein; 28 AA.

AC AAB44520;

DT 06-FEB-2001 (first entry)

XX Plant viral movement protein SEQ ID 56.

XX Plant viral movement protein; transport; transgenic plant;

KW viral resistance; cosuppression.

XX Triticum aestivum.

OS WO2000060088-A2.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US091110.

XX 07-APR-1999; 99US-0128092.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Krebbers E, Weng Z, Cahoon RE;

PI WPI; 2000-638467/61.

DR N-PSDB; AAC79375.

XX Novel viral movement polypeptides and polynucleotides useful in field
 PT of plant molecular biology, for producing transgenic plants, to prepare
 CC antibodies and in immunological screening of cDNA expression libraries

XX Claim 23; Page 62; 62pp; English.

XX Polynucleotide sequences AAC79348-C79375 encode plant viral movement
 CC proteins AAB44494-B44520. Some plant viruses have been shown to be able

CC to establish systemic infections via movement proteins that utilise
 CC existing plant pathways to traffic macromolecules to surrounding cells.
 CC Proteins such as those of the invention are similar to viral movement
 CC proteins that function in the transport of nucleic acids from cell to
 CC cell. The plant viral movement proteins are useful for obtaining a
 CC nucleic acid fragment encoding a viral movement protein. Polynucleotides
 CC encoding the plant viral movement proteins are useful for positive
 CC selection of a transformed cell. The proteins are useful in the field of
 CC plant molecular biology, and in the preparation of antibodies against the
 CC proteins. The proteins are also useful for isolating cDNAs and genes
 CC encoding homologous proteins from the same or other plant species, and to
 CC create transgenic plants in which the protein is presented at higher or
 CC lower levels than normal or in cell types or developmental stages in
 CC which they are not normally found. The proteins and nucleotide sequences
 CC may be used to control cosuppression and engineer plant virus
 CC resistance.
 XX

Seq Sequence 28 AA;

Query Match 93.5%; Score 29; DB 21; Length 28;
 Best Local Similarity 83.3%; Pred. No. 16;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 Db 1 MVHGKL 6
 :|||||

RESULT 12

AAB44518

ID AAB44518 standard; Protein; 69 AA.

XX AAB44518;

DT 06-FEB-2001 (first entry)

XX Plant viral movement protein SEQ ID 52.

XX Plant viral movement protein; transport; transgenic plant;
 KW viral resistance; cosuppression.

XX Oryza sativa.

XX WO2000060088-A2.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US091110.

XX 07-APR-1999; 99US-0128092.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Krebbers E, Weng Z, Cahoon RE;

XX WPI; 2000-638467/61.

DR N-PSDB; AAC79375.

XX Novel viral movement polypeptides and polynucleotides useful in field
 PT of plant molecular biology, for producing transgenic plants, to prepare
 CC antibodies and in immunological screening of cDNA expression libraries

XX Claim 23; Page 60; 62pp; English.

XX Polynucleotide sequences AAC79348-C79375 encode plant viral movement
 CC proteins AAB4494-B44520. Some plant viruses have been shown to be able
 CC to establish systemic infections via movement proteins that utilise
 CC existing plant pathways to traffic macromolecules to surrounding cells.
 CC Proteins such as those of the invention are similar to viral movement
 CC proteins that function in the transport of nucleic acids from cell to
 CC cell. The plant viral movement proteins are useful for obtaining a
 CC nucleic acid fragment encoding a viral movement protein. Polynucleotides

CC encoding the plant viral movement proteins are useful for positive
CC selection of a transformed cell. The proteins are useful in the field of
CC plant molecular biology, and in the preparation of antibodies against the
CC proteins. The proteins are also useful for isolating cDNAs and genes
CC encoding homologous proteins from the same or other plant species, and to
CC create transgenic plants in which the protein is presented at higher or
CC lower levels than normal or in cell types or developmental stages in
CC which they are not normally found. The proteins and nucleotide sequences
CC may be used to control cosuppression and engineer plant virus
CC resistance.

XX SQ Sequence 69 AA;

Query Match 93.5%; Score 29; DB 21; Length 69;

Best Local Similarity 83.3%; Pred. No. 42;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 1 MVHGKL 6

RESULT 13

AAB44507 ID AAB44507 standard; Protein: 145 AA.

XX AC AAB44507;

XX DT 06-FEB-2001 (first entry)

XX DE Plant viral movement protein SEQ ID 28.

XX KW Plant viral movement protein; transport; transgenic plant;

XX KW viral resistance; cosuppression.

XX OS Oryza sativa.

XX PN WO200060088-A2.

XX PD 12-OCT-2000.

XX PF 06-APR-2000; 2000WO-US09110.

XX PR 07-APR-1999; 99US-0128092.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Krebbers E, Weng Z, Cahoon RE;

XX WPI; 2000-638467/61.

XX DR N-PSDB; AAC79361.

XX Novel viral movement polypeptides and polynucleotides useful in field
XX of plant molecular biology, for producing transgenic plants, to prepare
XX antibodies and in immunological screening of cDNA expression libraries
XX -
XX Claim 10; Page 49; 62pp; English.

XX Polynucleotide sequences AAC79348-C79375 encode plant viral movement
XX proteins AAB4494-B44520. Some plant viruses have been shown to be able
XX to establish systemic infections via movement proteins that utilise
XX existing plant pathways to traffic macromolecules to surrounding cells.
XX Proteins such as those of the invention are similar to viral movement
XX proteins that function in the transport of nucleic acids from cell to
XX cell. The plant viral movement proteins are useful for obtaining a
XX nucleic acid fragment encoding a viral movement protein. Polynucleotides
XX encoding the plant viral movement proteins are useful for positive
XX selection of a transformed cell. The proteins are useful in the field of
XX plant molecular biology, and in the preparation of antibodies against the
XX proteins. The proteins are also useful for isolating cDNAs and genes
XX encoding homologous proteins from the same or other plant species, and to
XX create transgenic plants in which the protein is presented at higher or

CC lower levels than normal or in cell types or developmental stages in
CC which they are not normally found. The proteins and nucleotide sequences
CC may be used to control cosuppression and engineer plant virus
CC resistance.

XX SQ Sequence 145 AA;

Query Match 93.5%; Score 29; DB 21; Length 145;

Best Local Similarity 83.3%; Pred. No. 94;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 1 MVHGKL 6

RESULT 14

ABG25300 ID ABG25300 standard; Protein: 697 AA.

XX AC ABG25300;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #25291.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS89487.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 20; SEQ ID No 55659; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 697 AA;
Query Match 93.5%; Score 29; DB 22; Length 697;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
|||||
Db 379 LVHGKM 384

RESULT 15

ABR41404
ID ABR41404 standard; Protein; 751 AA.

XX ABR41404;

DT 02-JUN-2003 (first entry)

DE Human DITHP zinc finger transcriptional regulator.

KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW zinc finger; transcriptional regulator.

OS Homo sapiens.

XX WO200297031-A2.

PN 05-DEC-2002.

PD 27-MAR-2002; 2002WO-US10056.

PF 28-MAR-2001; 2001US-279619P.

PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280068P.

PR 16-MAY-2001; 2001US-291280P.

PR 17-MAY-2001; 2001US-291829P.

PR 17-MAY-2001; 2001US-291849P.

PR 19-JUN-2001; 2001US-299428P.

PR 20-JUN-2001; 2001US-299776P.

PR 20-JUN-2001; 2001US-300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RV, Urashka MB;

XX WPI: 2003-129518/12.

DR N-PSDB; ACC46344.

DR N-PSDB; ACC46344.

XX Novel human diagnostic and therapeutic polypeptide useful for
PT identifying test compound which specifically binds to a polypeptide
PT encoded by human diagnostic and therapeutic polynucleotide, and to
PT induce antibodies -

XX Claim 27; SEQ ID NO 939; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their
CC encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates
CC to polynucleotide sequences at least 90% identical to the dithp cDNA

CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods
CC of detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has zinc finger-
CC type transcriptional regulator activity.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 751 AA;

Query Match 93.5%; Score 29; DB 24; Length 751;
Best Local Similarity 83.3%; Pred. No. 5.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
|||||
Db 604 LVHGKM 609

RESULT 16

AAAY20541
ID AAAY20541 standard; Protein; 44 AA.

XX AAAY20541;

DT 22-JUL-1999 (first entry)

DE Human neurofilament-L mutant protein fragment 48.

KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
KW frameshift mutation; age-related disease; neurodegenerative disorder;
KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-W;
KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
KW high mobility group protein-C; neuroendocrine specific protein A.

XX Synthetic.

OS Homo sapiens.

XX WO9845322-A2.

PD 15-OCT-1998.

XX 02-APR-1998; 98WO-IB00705.

XX 10-APR-1997; 97US-0043163.

XX (UYUT-) RIJKSUNIV UTRECHT.

PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.

XX
PI Burbach JPH, Grosveld FG, Van Leeuwen FW;
XX
XX WPI; 1998-609901/51.
DR N-PSDB; AAX75758.
XX
XX
PT Diagnosing disease by detecting frameshift mutations in RNA or
PT corresponding protein mutations - used to diagnose cancer and
PT neurological diseases, particularly Alzheimer's disease, and also
PT for treatment and prevention with specific ribozymes or wild-type
PT RNA
XX
XX
PS Disclosure; Figure 7; 258pp; English.
XX
XX
CC This invention describes a novel method for the diagnosis of a disease
CC caused by, or associated with, an RNA molecule that has a frameshift
CC mutation. The method is used to diagnose age-related diseases, especially
CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC and many others listed) or susceptibility to these disorders. The method
CC allows a definitive diagnosis of Alzheimer's disease in living patients,
CC at an early stage. It is based on the observation that disease may be
CC caused by mutations in RNA rather than DNA. The invention describes the
CC use of neuronal system RNA molecules, specifically proteins including
CC beta-amyloid precursor protein (beta-APP), the microtubule associated
CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC neurofilament-F, presenilin 1, presenilin 2, glial fibrillary acidic
CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC protein-C (HMGP-C) and neuroendocrine specific protein A.
XX
XX
SQ Sequence 44 AA;

Query Match 90.3%; Score 28; DB 19; Length 44;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db : : : : :
7 IIHGKL 12

RESULT 17
ABG20244
ID ABG20244 standard; Protein; 81 AA.
XX
AC ABG20244;
XX
XX 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20235.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tanj YT;
PI
XX WPI; 2001-639362/73.
XX

DR N-PSDB; AAS84431.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 50603; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant products of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 81 AA;

Query Match 90.3%; Score 28; DB 22; Length 81;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db : : : : :
48 LVHGRL 53

RESULT 18
ABB82574
ID ABB82574 standard; Protein; 89 AA.
XX
AC ABB82574;
XX
XX 04-FEB-2003 (first entry)
XX
XX H. influenzae BASB223 or BASB224 polypeptide sequence.
XX
XX BASB223; BASB224; high molecular weight protein A; immunostimulant;
KW antibacterial; vaccine; medicament; bacterium.
XX
XX Haemophilus influenzae.
OS
XX WO200279237-A2.
PN
XX 10-OCT-2002.
PD
XX 12-MAR-2002; 2002WO-EP03210.
PF
XX 13-MAR-2001; 2001GB-0006155.
PR
XX 13-MAR-2001; 2001GB-0006156.
XX
XX (GLAX) GLAXOSMITHKLINE BIOLOGICALS SA.
XX
XX Thonnard J;
PI
XX WPI; 2003-040650/03.
DR
XX N-PSDB; ABV74998.
XX

PT New BASB223 and BASB224 polypeptides and polynucleotides, useful for
PT preparing a medicament for generating an immune response in an animal,
PT or as research reagents and materials for discovering treatments for
PT human diseases
XX
PS Claim 3; Page 71; 109pp; English.
XX
CC The invention relates to novel polypeptides BASB223 and BASB224, isolated
CC from non-typeable H. influenzae (NTHi). The polypeptides are related to
CC H. influenzae high molecular weight protein A (HmWA). Compositions
CC comprising the BASB223 and BASB224 polypeptides and polynucleotides are
CC useful for preparing a medicament for generating an immune response in an
CC animal. The BASB223 and BASB224 polypeptides and polynucleotides may be
CC used as research reagents and materials for discovery of treatments and
CC diagnosis for diseases, particularly human diseases, and as immunogens
CC for producing antibodies immunospecific the such polypeptides and
CC polynucleotides. The polynucleotides may be used in the discovery and
CC development of antibacterial compounds, in genetic immunization, in the
CC development of prophylactic or therapeutic agents for bacterial
CC infection, particularly non-typeable H. influenzae. The present sequence
CC represents the BASB223 or BASB224 polypeptide partial-sequence.
XX
SQ Sequence 89 AA;

Query Match 90.3%; Score 28; DB 24; Length 89;
Best Local Similarity 83.3%; Pred. No. 93;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 76 VVHGKL 81
:|||||

RESULT 19
AAU01706
ID AAU01706 standard; Protein; 92 AA.
XX
AC AAU01706;
XX
DT 29-AUG-2001 (first entry)
XX
DE Propetide of conopeptide Cj2, amino acid sequence.
XX
KW Gamma carboxyglutamate; neurological disorder; epilepsy; trauma;
KW hypoxia; anoxia; ischemia; stroke; brain; spinal cord; suffocation;
KW myocardial infarct; drowning; perinatal asphyxia; hypoglycaemia;
KW neurodegeneration; Alzheimer's disease; Huntington's disease;
KW senile dementia; Amyotrophic Lateral Sclerosis; multiple sclerosis;
KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;
KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;
KW seizure; chemical toxicity; addiction; dystonia; psychiatric disorder;
KW mood disorder; memory; ophthalmic; parasitic worm; conopeptide Cj2.
XX
OS Conus cinereus gubba.
XX
PN WO200118033-A1.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24816.
XX
PR 10-SEP-1999; 99US-0153034.
PR 21-JUL-2000; 2000US-0219673.
XX
PA (UTAH) UNIV UTAH RES FOUND.
PA (COGN-) COGNETIX INC.
XX
PI Olivera BM, McIntosh JM, Garrett JE, Walker CS, Watkins M;
PI Jones RM;
XX
DR WPI: 2001-273379/28.
DR N-PSDB; AAS02562.

PT New isolated gamma-carboxyglutamine containing peptide for treating or
PT preventing neurological and psychiatric disorders e.g. epilepsy,
PT Alzheimer's disease, migraine, chemical toxicity, dystonia, anxiety,
PT and depression
XX
PS Claim 5; Page 42; 102pp; English.
XX
CC The sequence represents the amino acid sequence of the propeptide of
CC gamma carboxyglutamate-containing conopeptide Cj2. The conopeptide
CC is used for treating or preventing disorders in which the pathophysiology
CC involves excess excitation of nerve cells by excitatory amino acids or
CC agonists of heterogeneous ionotropic glutamate receptors or heterogeneous
CC G protein coupled glutamate receptors. The disorders may be neurological
CC disorders, such as: (i) seizure associated with epilepsy; (ii) a
CC neurotoxic injury associated with hypoxia, anoxia, ischemia, stroke,
CC cerebrovascular accident, brain or spinal cord trauma, myocardial
CC infarct, physical trauma, drownings, suffocation, perinatal asphyxia, or
CC hypoglycaemic events; (iii) neurodegeneration associated with Alzheimer's
CC disease, Huntington's disease, senile dementia, Amyotrophic Lateral
CC Sclerosis, multiple sclerosis, Parkinson's disease, Down's Syndrome,
CC Korsakoff's disease, schizophrenia, AIDS (acquired immunodeficiency
CC syndrome) dementia from HIV infection, HIV infection, multi-infarct
CC dementia, Binswanger dementia and neuronal damage associated with
CC uncontrolled seizures; (iv) pain which is a migraine, acute pain, or
CC persistent pain; (v) chemical toxicity which is addiction, morphine,
CC opiate, opioid and barbiturate tolerance; and (vi) dystonia, urinary
CC incontinence, muscle relaxation or sleep disorder. The disorders may be
CC psychiatric disorders, such as, anxiety, major depression, manic-
CC depressive illness, obsessive-compulsive disorder, schizophrenia, or mood
CC disorders (bipolar disorder, unipolar depression, dysthymia, or seasonal
CC affective disorder). The conopeptide is also used to treat memory or
CC cognitive deficits, ophthalmic indications, or to control nematodes or
CC parasitic worms.
XX
SQ Sequence 92 AA;

Query Match 90.3%; Score 28; DB 22; Length 92;
Best Local Similarity 66.7%; Pred. No. 97;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 87 LIHGKI 92
:|||||

RESULT 20
ABG20246
ID ABG20246 standard; Protein; 106 AA.
XX
AC ABG20246;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #2037.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
XX N-PSDB; AAS84433.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 50605; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 106 AA;

Query Match 90.3%; Score 28; DB 22; Length 106;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db ||||:|
73 LVHGRL 78

RESULT 21
ABP06804
ID ABP06804 standard; Protein; 117 AA.
XX
AC ABP06804;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:13590.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX

PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN22556.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 13590; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27352 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 117 AA;

Query Match 90.3%; Score 28; DB 23; Length 117;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db ||||:|
49 LVHGRL 54

RESULT 22
ABB57434
ID ABB57434 standard; Protein; 173 AA.
XX
AC ABB57434;
XX
DT 15-MAR-2002 (first entry)
XX
DE Human secreted protein encoding polypeptide SEQ ID NO 80.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200183510-A1.
XX
PD 08-NOV-2001.

```

XX PF 26-APR-2001; 2001WO-US13318.
XX PA
XX XX
XX PI Komatsoulis G, Ruben SM, Rosen CA;
XX XX
XX PR 02-MAY-2000; 2000US-201194P.
XX PR 16-JUN-2000; 2000US-212142P.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Komatsoulis G, Ruben SM, Rosen CA;
XX XX
XX DR WPI; 2002-121886/16.
XX XX
XX PT An isolated nucleic acid molecule encoding a human secreted protein
XX PT (SP) is useful in preventing, treating or ameliorating a disorder e.g.,
XX PT Alzheimer's disease and cancers -
XX XX
XX PS Disclosure; Page 475; 496pp; English.
XX XX
XX CC The invention relates to novel genes (ABL01566-ABL01594) and proteins
XX CC (ABB57394-ABB57456) useful for preventing, treating or ameliorating
XX CC medical conditions e.g. by protein or gene therapy. The genes are
XX CC isolated from a range of human tissues disclosed in the specification.
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX CC and parasitic infections.
XX SQ Sequence 174 AA;

Query Match 90.3%; Score 28; DB 23; Length 173;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 120 LIHGKI 125
|:|:|:|:

RESULT 23
AAG30897
ID AAG30897 standard; Protein; 174 AA.
XX AC
XX XX
XX DT 17-OCT-2000 (first entry)
XX XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37018.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX XX
XX PN EP1033405-A2.
XX XX
XX PD 06-SEP-2000.
XX XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.

```

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-01593637.
PR	14-OCT-1999;	99US-01593638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-01595841.
PR	21-OCT-1999;	99US-01607675.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
PR	29-OCT-1999;	99US-0162144.

Query Match 90.3%; Score 28; DB 21; Length 178;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels

Qy 1 LVHGKL 6
|:| | | |
Db 7 LLHGKL 12

RESULT 25
ABB58792
ID ABB58792 standard; Protein: 280 AA.

AA
AC ABB58792;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 3168.

AA Drosophila; developmental biology; cell signalling; insecticide; insecticide;
KW pharmaceutical.
KW

AA
OS *Drosophila melanogaster*.

AX
PN
WO200171042-A2.

27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX	PI	Venter JC, Adams M, Li PWD, Mvers EW:
----	----	---------------------------------------

XX
DR WPI; 2001-656860/75.

DR WFI; Z001-636860/
DR N-PSDB; ABL02895.

xx
PT
PT
PT
PT

New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions -

PS Disclosure: SEO ID NO 3168: 21pp + Sequence Listing: English.

XX
CC The invention relates to an isolated nucleic acid detection reagent

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 280 AA;

Query Match	90.3%	Score 28;	DB 22;	Length 280;
Best Local Similarity	66.7%	Pred. No. 3.2e+02;		

Qy	1	LVHGKL	6
		::: ::	
Db	148	IHHGKL	153

RESULT 26

AAW58530
ID AAW58530 standard; Protein; 311 AA.

XX AAW58530:

XX
DT 01-SEP-1998 (first entry)

XX DE Human peroxisomal thioesterase.

XX Human; peroxisomal thioesterase; PXT; diagnosis; inflammatory disease;
KW fatty acid metabolism disorder; cancer; drug screening.
KW

XX
OS Homo sapiens.XX
PN
US5776753-A.

XX 07-JUL-1998.

XX 11-JUN-1997: 97US-0872784. PF

XX
PR 11-JUN-1997: 97US-0872784.

XX
PA (TNCY-) TNCYTE PHARM INC

XX
PT Corley NC: Hillman J.L. Shab P.

XX
DP WPT: 1008-308039/3A

DR WPI; 1998-398039/
DR N-PSDB: AAV31109

PT New nucleic acid encoding human peroxisomal thioesterase - useful
XX for diagnosis, treatment and prevention of fatty acid metabolism
PT disorders, cancer and inflammatory disease, and in drug screening
PT

XX
ps Claim 9. Fig 1. 30pp. English

xx The present sequence represents human peroxisomal thioesterase (PxTE).
cc The cDNA encoding PxTE was isolated from a human BRANT09 clone 2150905
cc cDNA library. PxTE and its agonists are used in treatment, diagnosis
cc and prevention of disorders of fatty acid metabolism e.g. neuronal
cc disorders, Zellweger or Down's syndrome, Refsum's disease, epilepsy,
cc Alzheimers disease, depression, tardive dyskinesia, multiple sclerosis,
cc Parkinson's disease, schizophrenia and Tourette's syndrome. Fragments of
cc PxTE DNA can also be used for diagnosis and monitoring (in hybridisation
cc or amplification methods), including determining a predisposition to
cc disease, for gene mapping and to detect PxTE-encoding nucleic acid,
cc particularly after amplification. PxTE can be used to raise antibodies
cc (Ab) used therapeutically, in competitive drug screens, as immunoassay
cc reagents for diagnosis and monitoring of diseases and for purification of
cc natural PxTE, and to screen for specific binding agents antagonists e.g.
cc specific Ab, antisense sequences or ribozymes which are used in treatment
cc of cancers and inflammatory diseases (e.g. adult respiratory distress
cc syndrome, asthma, diabetes mellitus, microbial and other infections).

CC Administration of proteins, agonist or antagonists, optionally combined
 CC with conventional pharmaceuticals providing a synergistic effect, is
 CC e.g. orally, by injection or topically.

SQ Sequence 311 AA;
 Query Match 90.3%; Score 28; DB 19; Length 311;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 ||||:|
 Db 278 LVHGRL 283

RESULT 27
 AAY08604
 ID AAY08604 standard; Protein; 311 AA.
 XX AC
 XX AAY08604;
 DT 05-AUG-1999 (first entry)
 XX DE
 XX Human PxTE protein.
 XX PxTE; peroxisomal thioesterase; treatment; disorder; biosynthesis;
 KW fatty acid metabolism; chain-termination; fatty acid; cancer; deficiency;
 KW inflammation; adrenoleukodystrophy Refsum's disease; mental retardation;
 KW disordered neuronal migration; hypotonia; tapetoretinal degeneration;
 KW sensorineural hearing loss; cystic change; kidney; skeletal change.
 XX OS
 XX Homo sapiens.
 XX PN
 XX US5911984-A.
 XX PD
 XX 15-JUN-1999.
 XX PF
 XX 19-JUN-1998; 98US-0100851.
 XX PR
 XX 11-JUN-1997; 97US-0872784.
 XX PR
 XX 19-JUN-1998; 98US-0100851.
 XX PA
 XX (INCY-) INCYTE PHARM INC.
 XX PI
 XX Corley NC, Hillman JL, Shah P;
 XX WPI: 1999-357169/30.
 XX DR
 XX N-PSDB; AAX77452.

XX Peroxisomal thioesterase enzymes useful for treating disorders
 associated with fatty acid metabolism

PS Claim 1; Fig 1A-C; 29pp; English.
 XX This invention describes a novel human peroxisomal thioesterase enzyme
 CC (PxTE) and its use to treat disorders associated with fatty acid
 CC metabolism. Thioesterases catalyse the chain-terminating step in the de
 CC novo biosynthesis of fatty acids. PxTE may be used to treat cancer,
 CC inflammation and disorders associated with fatty acid metabolism and the
 CC expression of PxTE, such as adrenoleukodystrophy Refsum's disease, and
 CC PxTE deficiency. Patients suffering from defective peroxisomal fatty
 CC acid metabolism exhibit disordered neuronal migration, hypotonia, mental
 CC retardation, tapetoretinal degeneration, sensorineural hearing loss,
 CC cystic changes in the kidney, skeletal changes and death.

SQ Sequence 311 AA;
 Query Match 90.3%; Score 28; DB 20; Length 311;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 ||||:|

Db 278 LVHGRL 283

RESULT 28
 AAG77835
 ID AAG77835 standard; protein; 311 AA.
 XX AC
 XX AAG77835;
 DT 03-JAN-2002 (first entry)
 XX DE
 XX Protein sequence of human peroxisomal thioesterase (PxTE).
 XX Human peroxisomal thioesterase; PxTE; immunogenic; cytostatic;
 KW neurotropic; neuroleptic; antiinflammatory; neuroprotective;
 KW cerebroprotective; Down's syndrome; gene therapy; cancer; inflammation;
 KW fatty acid metabolism disorders; neuronal disorders; Alzheimer's disease;
 KW amnesia; Huntington's disease; cerebrotendinous degeneration;
 KW hybridisation probe; adrenomyeloneuropathy; schizophrenia;
 KW adrenoleukodystrophy; Crohn's disease; adenocarcinoma; leukaemia;
 KW Addison's disease; adult respiratory distress syndrome.
 XX OS
 XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 309..311

XX FT /note="C-terminal SKL (ser-lys-leu) matrix targeting

XX FT sequence"

XX PN US01024786-A1.

XX PD 27-SEP-2001.

XX PF 18-JAN-2001; 2001US-0766366.

XX PR 11-JUN-1997; 97US-0872784.

XX PR 19-JUN-1998; 98US-0100851.

XX PR 09-MAR-1999; 99US-0265294.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Hillman JL, Shah P, Corley NC;

XX WPI: 2001-638498/73.

XX DR N-PSDB; AAG78821.

XX PT New human peroxisomal thioesterase polypeptide and polynucleotide,

XX PT useful in diagnosing, preventing and treating cancer, inflammation and

XX PT disorders associated with fatty acid metabolism

XX PS Claim 1; Fig 1; 33pp; English.

XX The protein sequence represents the human peroxisomal thioesterase
 CC enzyme (PxTE) of the invention. The invention also includes the
 CC polynucleotide sequence coding for this protein. Also the invention
 CC provides expression vectors, host cells, agonists, antibodies and
 CC antagonists. The invention also provides methods for treating disorders
 CC associated with expression of PxTE. The invention has the following
 CC activities of being cytostatic, antiinflammatory, neuroprotective,
 CC neurotropic, cerebroprotective and neuroleptic although no supporting data
 CC is given. The PxTE polypeptides are useful in diagnosing, preventing and
 CC treating cancer (e.g. adenocarcinoma or leukaemia), inflammation (e.g.
 CC colon and small intestine afflicted by Crohn's disease) and disorders
 CC associated with fatty acid metabolism, such as neuronal disorders
 CC including Alzheimer's disease, amnesia, Down syndrome, Huntington's
 CC disease, or schizophrenia by using the PxTE polypeptides in gene therapy.
 CC In one embodiment, an antagonist of PxTE may be administered to a subject
 CC to prevent or treat inflammation caused by conditions such as Addison's
 CC disease or adult respiratory distress syndrome. The PxTE polynucleotides
 CC can be used to detect and quantify gene expression in biopsied tissue. In
 CC the diagnosis of conditions, disorders or diseases associated with PxTE
 CC expression e.g. adrenoleukodystrophy, adrenomyeloneuropathy, or
 CC cerebrotendinous degeneration, to detect activation or induction of various

CC cancers, to generate hybridization probes useful for mapping naturally
 CC occurring genomic sequences, and for screening libraries in various drug
 CC screening techniques.

XX SQ Sequence 311 AA;
 Query Match 90.3%; Score 28; DB 22; Length 311;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
 ||||:|
 Db 278 LVHGRL 283

RESULT 29
 AAB73420
 ID AAB73420 standard; Protein; 311 AA.
 XX AC AAB73420;
 DT 02-JUL-2001 (first entry)
 XX DE Human peroxisomal thioesterase, PxTE.
 XX Human; peroxisomal thioesterase; PxTE; fatty acid biosynthesis;
 KW chain termination; detection; diagnosis; monitoring; hybridisation;
 KW inflammation; cancer; defective fatty acid metabolism; neuronal disorder;
 KW adrenoleukodystrophy; adrenomyeloneuropathy; cerebrotendinous degeneration;
 KW Zellweger syndrome; Refsum's disease; Alzheimer's disease; amnesia;
 KW amyotrophic lateral sclerosis; bipolar disorder; Huntington's disease;
 KW Down's syndrome; multiple sclerosis; neurofibromatosis; schizophrenia;
 KW Parkinson's disease; paranoid psychosis; Tourette's syndrome;
 KW drug screening.
 XX DE Human sapiens.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Region 309..311
 FT /note= "Peroxisomal matrix targeting sequence"
 XX US6210890-B1.
 PN 03-APR-2001.
 PD 09-MAR-1999; 99US-0265294.
 PF 11-JUN-1997; 97US-0872784.
 PR 19-JUN-1998; 98US-0100851.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Hillman JL, Shah P, Corley NC;
 PI WPI; 2001-289629/30.
 XX DR N-PSDB; AAF76400.
 XX Detecting human peroxisomal thioesterase (PxTE) in a biological sample,
 PT useful for diagnosing disorders associated with fatty acid metabolism,
 PT comprises hybridizing a PxTE polynucleotide complement with the sample
 PT nucleic acid -
 XX Claim 1; Fig 1A-C; 30pp; English.
 PS This sequence represents a human peroxisomal thioesterase, PxTE.
 XX PxTE, like other thioesterases, catalyses the chain-terminating
 CC step in the de novo biosynthesis of fatty acids. PxTE is expressed
 CC in neuronal, gastrointestinal and secretory tissues, and cells associated
 CC with inflammation and cancer. The invention relates to a method of
 CC detecting nucleic acids encoding PxTE involving the hybridisation of a
 CC PxTE nucleic acid probe to PxTE nucleic acids in a sample, and the
 CC subsequent detection of the hybridisation complex. The method is useful
 CC for the diagnosis of cancer, inflammation and disorders associated with

CC fatty acid metabolism, and may also be used to monitor the treatment of
 CC a patient with a disorder associated with PxTE expression. Disorders
 CC associated with defective fatty acid metabolism include neuronal
 CC disorders such as adrenoleukodystrophy, adrenomyeloneuropathy,
 CC cerebrotendinous degeneration (Zellweger syndrome), Refsum's disease,
 CC Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar
 CC disorder, Huntington's disease, Down's syndrome, multiple sclerosis,
 CC neurofibromatosis, Parkinson's disease, paranoid psychoses,
 CC schizophrenia, or Tourette's syndrome. In addition to their use as
 CC probes in the method of the invention, PxTE nucleic acids may also
 CC be used in chromosomal mapping, detecting polymorphisms associated
 CC with disease, in gene therapy, and in drug screening techniques. The
 CC PxTE protein may also be used in drug screening, and to raise
 CC antibodies.

XX SQ Sequence 311 AA;
 Query Match 90.3%; Score 28; DB 22; Length 311;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
 ||||:|
 Db 278 LVHGRL 283

RESULT 30
 ABG20249
 ID ABG20249 standard; Protein; 331 AA.
 XX AC ABG20249;
 DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #20240.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW Homo sapiens.
 OS WO200175067-A2.
 XX PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX DR N-PSDB; AAS84436.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID NO 50608; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 331 AA;

Query Match 90.3%; Score 28; DB 22; Length 331;
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
 |||||
 Db 298 LVHGRL 303

RESULT 31

AAB36830

ID AAB36830 standard; Protein; 341 AA.

XX AC AAB36830;

XX DT 19-FEB-2001 (first entry)

XX DE 37kD protein.

XX KW 37kD protein; cattle; infection.

XX OS Babesia divergens.

XX PN EP1050541-A1.

XX PD 08-NOV-2000.

XX PF 25-APR-2000; 2000EP-0201485.

XX PR 29-APR-1999; 99EP-0201322.

XX PA (ALKU) AK20 NOBEL NV.

XX PI Schettlers TPM, Carcy B, Gorenflot A, Precigout E, Vallat A;

XX DR WPI; 2000-681240/67.

XX DR N-PSDB; AAC60471.

XX PT New nucleic acid encoding protein of the 37 kD protein family of

XX PS Babesia divergens, useful in e.g. protective vaccines,

XX Claim 5; Page 22-23; 48pp; English.

XX CC The present invention relates to the 37 kD protein family or its

XX CC immunogenic fragment of Babesia divergens. Derived vectors or host

XX CC cells, are useful in vaccines to protect cattle against B. divergens

XX CC infection, or at least its symptoms. Vaccines based on the invention

XX CC protect against both homologous and heterologous strains of B.

XX CC divergens.

XX SQ Sequence 341 AA;

Query Match 90.3%; Score 28; DB 21; Length 341;
 Best Local Similarity 66.7%; Pred. No. 4e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
 ::|||

Db 266 MIHGKL 271

RESULT 32

AAG64743

ID AAG64743 standard; Protein; 392 AA.

XX AC AAG64743;

XX DT 25-SEP-2001 (first entry)

XX DE Human sodium neurotransmitter cotransporter protein 8,43.

XX KW Human; sodium neurotransmitter cotransporter protein 8,43; cytostatic;

XX KW virucidal; immunomodulatory; antiinflammatory; haemostatic;

XX KW malignant tumour; haemopathy; HIV infection; immunological disease;

XX KW inflammatory condition.

XX OS Homo sapiens.

XX PN WO200153496-A1.

XX PD 26-JUL-2001.

XX PF 15-JAN-2001; 2001WO-CN00020.

XX PR 21-JAN-2000; 2000CN-0111469.

XX PA (BIOD-) BIODOR GENE TECHNOLOGY LTD SHANGHAI.

XX PI Mao Y, Xie Y;

XX DR WPI; 2001-465376/50.

XX DR N-PSDB; AAH73714.

XX PT Human sodium neurotransmitter cotransporter protein 8,43 and encoded

XX PT polynucleotide, applicable in diagnosis and treatment of malignant

XX PT tumor, hemopathy, HIV infection, immunological diseases and various

XX PT inflammations

XX PS Claim 1; Page 33-34; 40pp; Chinese.

XX CC This invention relates to human sodium neurotransmitter cotransporter

XX CC protein 8,43 and the cDNA encoding it. Included in the invention is a

XX CC vector containing the cDNA, a host cell transformed with the vector and

XX CC an antibody targeting the protein. The protein, polynucleotide, antibody,

XX CC or antagonist can be used to achieve cytostatic; virucidal;

XX CC immunomodulatory; antiinflammatory or haemostatic activity. They can be

XX CC used in the diagnosis and treatment of malignant tumours, haemopathy, HIV

XX CC infection, immunological diseases and various inflammatory conditions.

XX CC The present sequence represents the human sodium neurotransmitter

XX CC cotransporter protein 8,43.

XX SQ Sequence 392 AA;

Query Match 90.3%; Score 28; DB 22; Length 392;

Best Local Similarity 66.7%; Pred. No. 4.7e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
 |::|||

Db 338 LIHGKI 343

RESULT 33

AAY35719

ID AAY35719 standard; Protein; 554 AA.

XX AC AAY35719;

XX DT 13-SEP-1999 (first entry)

XX DE C. pneumoniae protein involved in metabolism of nucleic acids.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

XX WO927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 1423-1424; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 554 AA;

Query Match 90.3%; Score 28; DB 20; Length 554;
 Best Local Similarity 66.7%; Pred. No. 6.8e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 77 IHHGKL 82

RESULT 34

ABB49246

ID ABB49246 standard; Protein; 682 AA.

XX ABB49246;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #1950.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

XX

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

Buchrieser C, Prangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 Dusurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 Rose M, Voss H;

WPI; 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 and prevention of Listeria and related bacterial infections, and
 related polypeptides

Claim 6; SEQ ID No 1951; 192pp; French.

The present invention relates to the genome sequence of Listeria
 monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 it are useful for selecting probes and primers for detecting genes in L.
 monocytogenes and related organisms, and for studying genetic
 polymorphisms and other genomes. The present sequence is a protein
 encoded by the genome sequence of the present invention. Proteins
 expressed from the genome sequence are useful for raising specific
 antibodies, identification of L. monocytogenes and related organisms, and
 for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 B12. The genome sequence and proteins encoded by it are also useful for
 selecting compounds that regulate gene expression and cell replication
 and modulate L. monocytogenes-related diseases. In addition, the genome
 sequence and proteins encoded by it are useful in pharmaceutical and
 vaccines compositions for the treatment or prevention of infections by L.
 monocytogenes and related organisms.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 682 AA;

Query Match 90.3%; Score 28; DB 23; Length 682;

Best Local Similarity 83.3%; Pred. No. 8.9e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 506 LMHGKL 511

RESULT 35

AAO21215

ID AAO21215 standard; Protein; 729 AA.

XX AAO21215;

XX 19-JUL-2002 (first entry)

XX Protein sequence identified by SwissProt Accession No: Q08469.

XX Antiparkinsonian; nootropic; neuroprotective; tranquiliser; vulnery;
 KW cerebroprotective; human sodium-dependent neurotransmitter transporter;
 KW hybridising; central; peripheral nervous system disease; brain injury;
 KW cerebrovascular disease; Parkinson's disease; corticobasal degeneration;
 KW motor neuron disease; dementia; multiple sclerosis; post-stroke;
 KW traumatic brain injury; stroke; post-traumatic brain injury;
 KW small-vessel cerebrovascular disease; Alzheimer's disease; gene therapy.

OS Homo sapiens.

XX WO200229048-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-EP11440.

XX 05-OCT-2000; 2000US-237689P.
 XX (FARB) BAYER AG.
 PA Kohler RH;
 PI
 XX WPI; 2002-426112/45.
 XX
 XX New human sodium-dependent neurotransmitter transporter polypeptide,
 PT the regulation of which is useful for treating central or peripheral
 PT nervous system diseases e.g. brain injuries, Parkinson's disease, and
 PT dementia
 XX
 XX Disclosure; Fig 3; 98pp; English.
 PS
 XX The invention relates to a purified human sodium-dependent
 XX neurotransmitter transporter polypeptide comprising a sequence of 259
 CC amino acids, given in the specification. A nucleic acid encoding the
 CC protein is useful for detecting a polynucleotide encoding the protein in
 CC a biological sample by hybridising the protein-encoding polynucleotide to
 CC a nucleic acid material of a biological sample to form a hybridisation
 CC complex, and detecting the hybridisation complex. The protein of the
 CC invention and its encoding gene are useful for screening for agents which
 CC decrease the activity of sodium-dependent neurotransmitter transporter
 CC polypeptide by contacting the test compound with the protein or its gene
 CC and detecting binding of the test compound to the protein or its gene. A
 CC vector comprising the polynucleotide of the invention or a reagent that
 CC modulates the activity of the protein or its gene is useful for the
 CC preparation of a medicament for modulating the activity of sodium-
 CC dependent neurotransmitter transporter in a disease. The reagent is
 CC useful for treating diseases such as central or peripheral nervous system
 CC disease, where the central nervous system disorders are selected from
 CC brain injuries, cerebrovascular diseases and their consequences,
 CC Parkinson's disease, corticobasal degeneration, motor neuron disease,
 CC dementia, including multiple sclerosis, traumatic brain injury, stroke,
 CC post-stroke, post-traumatic brain injury, small-vessel cerebrovascular
 CC disease, and Alzheimer's disease. The polynucleotide of the invention can
 CC be used in gene therapy. This sequence represents the protein sequence
 CC identified by SwissProt Accession No: Q08469 relating to the sodium-
 CC dependent neurotransmitter transporter protein of the invention.
 XX
 XX Sequence 729 AA;
 SQ
 Query Match 90.3%; Score 28; DB 23; Length 729;
 Best Local Similarity 66.7%; Pred No. 9.1e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db I:III:
 676 LIHGKI 681
 RESULT 36
 AAG67159
 ID AAG67159 standard; Protein; 730 AA.
 XX
 AC AAG67159;
 XX
 DT 13-NOV-2001 (first entry)
 XX
 XX Amino acid sequence of a human 579 transporter polypeptide.
 DE
 DE Human; transporter; 20685; 579; 17114; 23821; 33894; 32613;
 KW vesicular monoamine transporter; neurotransmitter-symporter;
 KW ABC transporter; sulfate transporter; neurological disorder;
 KW central nervous system disorder; Parkinson's disease; depression; pain;
 KW infectious disease; cell proliferative disorder; cancer; blood disorder;
 KW immune disorder; inflammatory disorder; spleen disorder; lung disorder;
 KW Hodgkin's disease; Niemann-Pick disease; chronic bronchitis; ischemia;
 KW colon disorder; cirrhosis; uterus disorder; endometrial disorder;
 KW endometrial stromal tumour; brain disorder; T-cell disorder; anemia;
 KW Sjogren syndrome; skin disorder; lupus erythematosus; heart disorder;

KW haematopoietic stem cell; Alzheimer's disease; myocardial infarction;
 KW blood vessel; Kawasaki syndrome; red cell disorder; thymus disorder;
 KW B-cell disorder; kidney disorder; glomerulonephritis; breast disorder;
 KW testis disorder; thyroid disorder; Graves disease; pancreatitis;
 KW skeletal muscle disorder; tumour; pancreas disorder;
 KW small intestine disorder; celiac sprue.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Location/Qualifiers
 FH 61..659
 FT Domain /note= "sodium neurotransmitter-symporter domain"
 FT Domain 70..87
 FT Domain /note= "transmembrane domain"
 FT Domain 98..117
 FT Domain /note= "transmembrane domain"
 FT Domain 140..164
 FT Domain /note= "transmembrane domain"
 FT Domain 228..244
 FT Domain /note= "transmembrane domain"
 FT Domain 253..275
 FT Domain /note= "transmembrane domain"
 FT Domain 306..323
 FT Domain /note= "transmembrane domain"
 FT Domain 334..358
 FT Domain /note= "transmembrane domain"
 FT Domain 458..479
 FT Domain /note= "transmembrane domain"
 FT Domain 496..513
 FT Domain /note= "transmembrane domain"
 FT Domain 527..550
 FT Domain /note= "transmembrane domain"
 FT Domain 575..594
 FT Domain /note= "transmembrane domain"
 FT Domain 617..639
 FT Domain /note= "transmembrane domain"
 XX
 XX WO200164875-A2.
 XX 07-SEP-2001.
 XX
 XX 28-FEB-2001; 2001WO-US06374.
 XX
 XX 29-FEB-2000; 2000US-0185906.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Glucksmann MA;
 XX
 XX WPI; 2001-550178/61.
 XX N-PSDB; AAH75186.
 XX
 XX Novel human transporter polypeptides useful for treating and diagnosing
 PT Parkinson's disease, Hodgkin disease, glomerulonephritis, myocardial
 PT infarction, Grave's disease, Alzheimer's disease, anemia, asthma and
 PT tumours -
 XX
 XX Claim 9; Fig 9A-C; 259pp; English.
 PS
 XX The present sequence represents a human transporter polypeptide. The
 CC specification describes 20685, 579, 17114, 23821, 33894 or 32613 human
 CC transporter polypeptides. The 20685 transporter is similar to vesicular
 CC monoamine transporters. The 579 transporter is similar to
 CC neurotransmitter-symporters. The 17114 transporter is similar to ABC
 CC transporters. The 32613 transporter is similar to sulfate transporters.
 CC The transporter polypeptides and polynucleotides are useful for treating
 CC and diagnosing neurological and central nervous system disorders (e.g.
 CC Parkinson's disease, depression, pain), infectious disease, cell
 CC proliferative disorders (e.g., cancer), blood disorders, and immune and
 CC inflammatory disorders. They are also useful for treating and
 CC diagnosing disorders involving the spleen (e.g., Hodgkin disease,
 CC Niemann-Pick disease), lung (e.g., chronic bronchitis), colon
 CC (cirrhosis), uterus and endometrium (e.g., endometrial stromal tumours),

CC brain (e.g., ischemia), T-cells (e.g., Sjogren syndrome), skin (lupus erythematosus), haematopoietic stem cells (e.g., Alzheimer's disease), heart (e.g., myocardial infarction), blood vessels (e.g., Kawasaki syndrome), red cells (e.g., anemias), disorders involving thymus, B-cells, kidney (e.g., glomerulonephritis), disorders involving breast, C-cells, epidiymis, prostate, thyroid (e.g., Graves disease), disorders involving skeletal muscle (e.g., tumour), pancreas (e.g., pancreatitis), small intestine (e.g., celiac sprue), disorders related to reduced platelet number and ovary.

SQ Sequence 730 AA;

Query Match 90.3%; Score 28; DB 22; Length 730;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LVHGKL 6
Db 676 LIHGKI 681
I:|:|:|:

RESULT 37
AA72908

ID AAY72908 standard; Protein; 730 AA.

XX AC AAY72908;

XX DT 13-JUN-2001 (first entry)

XX DE Human NTF7 protein.

XX KW Human; NTF7 protein; therapy; psychiatric disorder; anxiety; depression; schizophrenia; phobia; panic disorder; obsessive compulsive disorder; Parkinson's disease; central nervous system disorder; cerebroprotective; neurological disorder; stroke; pain; neuropathic pain; sleep disorder; tranquilliser; analgesic; neuroleptic; vaccine.

XX OS Homo sapiens.

XX PN WO200119854-A2.

XX PD 22-MAR-2001.

XX PF 14-SEP-2000; 2000WO-GB03541.

XX PR 15-SEP-1999; 99GB-0021833.

XX PR 17-MAR-2000; 2000GB-0006545.

XX PA (SMIK) SMITHKLINE BEECHAM PLC.

XX PI Hill J, Duckworth DM, Farmer M, Pangalos M;

XX DR WPI: 2001-244780/25.

XX DR N-PSDB; AAD03039.

XX PT Novel NTF7 polypeptide useful for treating anxiety, depression, schizophrenia, phobia, Parkinson's disease, stroke, pain, and psychiatric, panic, neurological, central nervous system, obsessive compulsive and sleep disorders

XX PS Claim 3; Page 20; 27pp; English.

XX CC The present sequence is human NTF7 protein encoded by a cDNA. NTF7 is thought to be a member of the neurotransmitter family of polypeptides. NTF7 sequences are useful for treating psychiatric disorders, anxiety, depression, schizophrenia, phobias, panic disorder, obsessive compulsive disorder, Parkinson's disease, central nervous system disorders, stroke, neurological disorders, pain, neuropathic pain, sleep disorders, and diseases in which neurotransmitters are implicated. NTF7 sequences are useful for screening antagonists and agonists of NTF7 and as vaccines for inducing immunological response in a mammal. NTF7 sequences and its antibodies are useful to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in

CC cells. NTF7 is useful in conventional low capacity screening methods and also in high-throughput screening (HTS) formats and is useful for identifying membrane bound or soluble receptors. NTF7 antibody is useful to isolate or identify clones expressing NTF7 or to purify NTF7 by affinity chromatography. NTF7 is useful as diagnostic reagent for detecting mutations in the associated gene, and for chromosome localisation studies and tissue expression studies.

SQ Sequence 730 AA;

Query Match 90.3%; Score 28; DB 22; Length 730;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LVHGKL 6
Db 676 LIHGKI 681
I:|:|:|:

RESULT 38
ABB77168

ID ABB77168 standard; Protein; 730 AA.

XX AC ABB77168;

XX DT 15-JUL-2002 (first entry)

XX DE Human GABA transporter protein.

XX KW Human; GABA; transporter protein; drug screening; therapeutic target.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX FH Region 19..21 /label= Protein_kinase_C_phosphorylation_site

XX FT Region 19..22 /label= Casein_kinase_II_phosphorylation_site

XX FT Region 25..28 /label= Casein_kinase_II_phosphorylation_site

XX FT Region 49..52 /label= Casein_kinase_II_phosphorylation_site

XX FT Region 55..58 /label= Casein_kinase_II_phosphorylation_site

XX FT Region 85..99 /label= Sodium:neurotransmitter_symporter_family_signature_1

XX FT Region 126..129 /label= CAMP_and_cGMP-dependant_protein_kinase_phosphorylation_site

XX FT Region 137..139 /label= Protein_kinase_C_phosphorylation_site

XX FT Modified-site 142..147 /note= "Gly is N-myristoylated"

XX FT Modified-site 187..190 /note= "Asn is N-glycosylated"

XX FT Region 191..194 /label= Casein_kinase_II_phosphorylation_site

XX FT Modified-site 213..216 /note= "Asn is N-glycosylated"

XX FT Region 217..220 /label= Casein_Kinase_II_phosphorylation_site

XX FT Modified-site 246..251 /note= "Gly is N-myristoylated"

XX FT Region 250..252 /label= Protein_kinase_C_phosphorylation_site

XX FT Modified-site 276..279 /note= "Asn is N-glycosylated"

XX FT Modified-site 277..282 /note= "Gly is N-myristoylated"

XX FT Region 287..289 /label= Protein_kinase_C_phosphorylation_site

XX FT Modified-site 311..316

FT /note= "Gly is N-myristoylated"
FT 315..320
FT /note= "Gly is N-myristoylated"
FT 383..386
FT /note= "Asn is N-glycosylated"
FT 394..397
FT /note= "Asn is N-glycosylated"
FT 399..402
FT /label= Casein_kinase_II_phosphorylation_site
FT 425..427
FT /label= Protein_kinase_C_phosphorylation_site
FT 438..443
FT /note= "Gly is N-myristoylated"
FT 473..478
FT /note= "Gly is N-myristoylated"
FT 477..482
FT /note= "Gly is N-myristoylated"
FT 489..491
FT /label= Protein_kinase_C_phosphorylation_site
FT 522..525
FT /label= Casein_kinase_II_phosphorylation_site
FT 606..614
FT /label= Tyrosine_kinase_phosphorylation_site
FT 618..623
FT /note= "Gly is N-myristoylated"
FT 650..655
FT /note= "Gly is N-myristoylated"
FT 656..658
FT /label= Protein_kinase_C_phosphorylation_site
FT 696..699
FT /label= cAMP- and cGMP-dependent_protein_kinase_phosphorylation_site
FT
FT
PN US2002031800-A1.
XX
XX 14-MAR-2002.
PD
XX
XX 21-DEC-2000; 2000US-07411149.
XX
XX 05-SEP-2000; 2000US-229529P.
XX
XX (LI2Z/) LI Z.
PA (CHAT/) CHATURVEDI K.
PA (ZHUS/) ZHU S.
PA (WOOD/) WOODAGE T.
PA (GUEG/) GUEGLER K.
PA (WEBB/) WEBSTER M.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
XX Li Z. Chaturvedi K, Zhu S, Woodage T, Guegler K, Webster M;
PI Di Francesco V, Beasley EM;
PI
XX WPI; 2002-361179/39.
DR N-PSDB; ABL55856, ABL55857.
XX
XX New peptides and nucleic acid sequences related to the GABA transporter subfamily, useful in developing diagnostic compositions, as well as in drug screening, particularly as models for the development of human therapeutic targets -
PT
XX
XX Claim 1; Fig 2; 69pp; English.
PS
XX The sequence represents a novel human transporter protein of the GABA transporter subfamily. The invention relates to a novel isolated human peptide of the GABA transporter subfamily. The peptide and nucleic acids are useful in developing human therapeutics and diagnostic compositions. These are also useful in drug screening, particularly as models for the development of human therapeutic targets.
CC
XX Sequence 730 AA;
SQ

Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVHGKL 6
DB 676 LIHGKI 681
RESULT 39
AAEL1404
ID AAE14404 standard; Protein; 730 AA.
XX
AC AAE14404;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human neurotransmitter transporter, NTT-2.
XX
KW Human; neurotransmitter transporter; NTT-2; transport disorder;
KW neurological disorder; psychiatric disorder; diabetes mellitus;
KW amyotrophic lateral sclerosis; Parkinson's disease; prostate cancer;
KW cardiac disorder; angina; Alzheimer's disease; amnesia; epilepsy;
KW sickle cell anaemia; stroke; Huntington's disease; psychiatric disorder;
KW acute stress disorder; anorexia nervosa; transgenic animal;
KW gene therapy; sodium:neurotransmitter symporter family; SNF.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 100..117
FT /label= Transmembrane_domain
FT Domain 335..359
FT /label= Transmembrane_domain
FT Domain 460..483
FT /label= Transmembrane_domain
FT Domain 494..512
FT /label= Transmembrane_domain
FT Domain 572..589
FT /label= Transmembrane_domain
FT Domain 619..640
FT /label= Transmembrane_domain
XX
XX WO200190148-A2.
XX
XX 29-NOV-2001.
XX
XX 17-MAY-2001; 2001WO-US16283.
XX
XX 19-MAY-2000; 2000US-205518P.
XX 22-JUN-2000; 2000US-213956P.
XX 28-JUN-2000; 2000US-215105P.
XX 14-JUL-2000; 2000US-218947P.
XX 27-JUL-2000; 2000US-228448P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Sanjanwala MS, Walia NK, Tribouley CM, Yue H, Gandhi AR, Ding L;
PI Rao MG, Lal P, Baughn MR, Hafalia A, Elliott VS, Patterson C;
PI Rankumar J;
XX
XX WPI; 2002-097640/13.
DR N-PSDB; AAD23974.
XX
XX Novel human neurotransmitter transporter polypeptides and
PT polynucleotides for diagnosing, preventing or treating transport,
CC neurological and psychiatric disorders and for identifying modulators
CC of therapeutic use -
XX
XX Claim 1; Page 111-113; 123pp; English.
PS
XX The present sequence is human neurotransmitter transporter (NTT)-2
CC (IncYTE ID No: 6881669CD1). The NTT-2 polypeptide contains
CC sodium:neurotransmitter symporter family (SNF) signature

CC sequences. The NTP polypeptide and polynucleotide
CC are useful for diagnosis, treatment and prevention of transport,
CC neurological and psychiatric disorders. Transport disorders
CC include akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia,
CC cystic fibrosis, Becker's muscular dystrophy, diabetes mellitus,
CC diabetes insipidus, myasthenia gravis, myocarditis, Parkinson's
CC disease, prostate cancer; cardiac disorders associated with transport
CC include angina, bradyarrhythmia, dermatomyositis, polymyositis;
CC neurological disorders associated with transport include Alzheimer's
CC disease, amnesia, bipolar disorder, dementia, depression, epilepsy,
CC Tourette's disorder, schizophrenia, and other disorders associated with
CC transport include neurofibromatosis, sickle cell anaemia, Wilson's
CC disease, cataracts, infertility, hyperglycaemia, hypoglycaemia, Graves'
CC disease, goitre, Cushing's disease, hypercholesterolaemia and
CC cystinuria. Neurological disorders treatable include epilepsy, stroke,
CC Huntington's disease, dementia, and other extrapyramidal disorder,
CC motor neuron disorders, prion disease including kuru, metabolic disease
CC of the nervous system, and other developmental disorders of the central
CC nervous system, neuromuscular disorders, metabolic, endocrine and toxic
CC myopathies, periodic paralysis, mental disorders including mood and
CC anxiety. Psychiatric disorders include acute stress disorder, alcohol
CC dependence, anorexia nervosa, anxiety, obsessive-compulsive disorder,
CC panic disorder and sleep disorder. The polynucleotide is useful for
CC creating knockin humanised animals or transgenic animals to model
CC human disease and to detect and quantify gene expression in biopsied
CC tissues in which expression of NTP is correlated with disease. The
CC polynucleotide is also useful for generating hybridisation probes useful
CC in mapping the naturally occurring genomic sequence and oligonucleotide
CC primers derived from it are useful to detect single nucleotide
CC polymorphisms. NTP, its fragments and antibodies are useful as elements
CC on a microarray which is useful to monitor or measure protein-protein
CC interactions, drug-target interactions and gene expression profiles.
CC Sequences of the NTP polypeptide are used to analyse the proteome of a
CC tissue or cell type. The polypeptide of the invention is also useful for
CC screening its agonist, antagonist, modulator or a compound that binds
CC to it.

XX
SQ Sequence 730 AA;
Query Match 90.3%; Score 28; DB 23; Length 730;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 676 LIHGKI 681
:|||||

RESULT 40
ABJ25504
ID ABJ25504 standard; Protein; 759 AA.

XX AC ABJ25504;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene protein #162.

XX DE Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.

XX OS Aspergillus fumigatus.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US13142.

XX PR 23-APR-2001; 2001US-285697P.

XX PR 27-APR-2001; 2001US-287066P.

XX PR 05-JUN-2001; 2001US-295890P.

XX PR 09-JUL-2001; 2001US-303899P.

PR 31-AUG-2001; 2001US-316362P.
XX (ELIT-) ELITRA PHARM INC.
XX JIANG B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PI WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer
XX Disclosure; Page -; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.

XX
SQ Sequence 759 AA;
Query Match 90.3%; Score 28; DB 24; Length 759;
Best Local Similarity 83.3%; Pred. No. 9.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 43 VVHGKL 48
:|||||

RESULT 41
AAR42962
ID AAR42962 standard; Protein; 860 AA.

XX AC AAR42962;
XX DT 10-MAY-1994 (first entry)
XX DE Bovine MTP.
XX KW MTP; microsomal triglyceride transfer protein; obesity;
KW atherosclerosis; detection; inhibitor; prevention; stabilisation;
KW regression; lipid; serum cholesterol; SC; triglyceride; TG;
KW phosphatidyl choline; PC; cholesterol ester; CE;
KW pancreatitis; hyperglycemia.
XX OS Bos taurus.
XX PN AU9334064-A.

PD 09-SEP-1993.
 XX
 PF 09-MAR-1993; 93AU-0034064.
 XX
 PR 06-MAR-1992; 92US-0847503.
 XX
 PA (SQUI) SQUIBB & SONS INC E R.
 PI (WETT/) WETTERAU J R.
 XX
 PI Gregg RE, Sharp DY, Wetterau JR;
 XX
 DR WPI; 1993-337295/43.
 DR N-PSDB; AAQ50429.
 XX
 PT Nucleic acid sequence encoding sub-unit of WTP - used to express
 PT polypeptide for therapeutic use e.g. in reduction of obesity
 PT atherosclerosis etc.
 XX
 PS Claim 24; Page 25; 112pp; English.
 XX
 CC The sequences given in AAQ50429-33 may be used for the detection
 CC of WTP or WTP inhibitors.
 CC Bovine and human WTP are useful for the prevention, stabilisation or
 CC regression of atherosclerosis or for lowering the level of serum
 CC lipids such as serum cholesterol, triglyceride, phosphatidyl choline
 CC or cholesteryl ester in a mammalian species.
 CC Diseases associated or affected by serum lipid levels, such as
 CC pancreatitis, hyperglycemia, obesity, etc., may be treated.
 XX
 SQ Sequence 860 AA;
 Query Match 90.3%; Score 28; DB 14; Length 860;
 Best Local Similarity 66.7%; Pred. No. 1.1e+03;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LVHGKL 6
 DB 93 LIHGKI 98
 RESULT 42
 ABJ26104
 ID ABJ26104 standard; Protein; 869 AA.
 XX
 AC ABJ26104;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene protein #762.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response.
 XX
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US13142.
 XX
 PR 23-APR-2001; 2001US-285697P.
 PR 27-APR-2001; 2001US-287066P.
 PR 05-JUN-2001; 2001US-295890P.
 PR 09-JUL-2001; 2001US-303899P.
 PR 31-AUG-2001; 2001US-316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX

PT New purified or isolated nucleic acids of essential genes of
 PT Aspergillus fumigatus, useful for treating or preventing infections by
 PT A. fumigatus, or for treating a non-infectious disease in a subject
 XX e.g. cancer
 PS Disclosure; Page -; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This sequence represents a protein of one of the essential genes
 CC of Aspergillus fumigatus of the invention.
 XX
 SQ Sequence 869 AA;
 Query Match 90.3%; Score 28; DB 24; Length 869;
 Best Local Similarity 83.3%; Pred. No. 1.1e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 LVHGKL 6
 DB 175 VVHGKL 180
 RESULT 43
 AAE18115
 ID AAE18115 standard; Protein; 914 AA.
 XX
 AC AAE18115;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DE Human MDM2 binding protein (MTBP).
 XX
 KW Human; MDM2 binding protein; MTBP; cytostatic; growth pathway; tumour;
 KW cell cycle; G1 phase.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 9..11 /note= "Encoded by ATCTGG"
 XX
 PN WO200204601-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 12-JUL-2001; 2001WO-US22053.
 XX
 PR 12-JUL-2000; 2000US-217835P.
 PR 05-SEP-2000; 2000US-230894P.
 XX

PA (PHIL-) PHILADELPHIA HEALTH & EDUCATION CORP.

PI Boyd MT, Haines DS;

DR WPI; 2002-171711/22.

DR N-PSDB; AAD29106.

XX Novel MDM2 binding protein and nucleic acid sequence encoding the
PT protein, useful for inducing an arrest in cell cycle, and hence for
PT treating tumor

XX Claim 6; Page 25-29; 34pp; English.

XX The invention relates to an isolated MDM2 binding protein (MTBP) and
CC nucleic acid sequence encoding the protein. MTBP is involved in MDM2
CC growth regulatory pathways. A composition comprising MTBP is useful for
CC modulating the G1 phase of cell cycle. MTBP is useful for raising
CC antibodies targeted to the mammalian MTBP polypeptide. The composition
CC is also useful for suppressing tumours. The present sequence is human
CC MTBP protein.

XX Sequence 914 AA;

Query Match 90.3%; Score 28; DB 23; Length 914;

Best Local Similarity 66.7%; Pred. No. 1.2e+03;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKLL 6

I:||||

Db 433 LIHGKM 438

RESULT 44

ABB64494

ID ABB64494 standard; Protein; 2248 AA.

XX AC ABB64494;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 20274.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08597.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions

XX Disclosure; SEQ ID NO 20274; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2248 AA;

Query Match 90.3%; Score 28; DB 22; Length 2248;

Best Local Similarity 83.3%; Pred. No. 3.1e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKLL 6

I:||||

Db 893 LVHGRL 898

RESULT 45

ABG49213

ID ABG49213 standard; Peptide; 46 AA.

XX AC ABG49213;

XX 25-FEB-2003 (first entry)

XX Human liver peptide, SEQ ID No 27861.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00664.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234587.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver

XX Claim 27; SEQ ID No 27861; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (I) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABB47348-ABG59930 represent
CC human liver single exon encoded peptides of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

xx
SQ

Sequence 46 AA;

Query Match 87.1%; Score 27; DB 22; Length 46;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
::|||
Db 12 VIHGKL 17

Search completed: October 14, 2003, 13:14:19
Job time : 46.3636 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 15.8182 Seconds
(without alignments)
16.049 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29
Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/FctUS_COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	31	100.0	391	US-09-328-352-4861	Sequence 4861, Ap
2	29	93.5	684	US-09-328-352-6053	Sequence 6053, Ap
3	29	93.5	1005	US-09-328-352-4877	Sequence 4877, Ap
4	28	90.3	311	US-08-872-784-1	Sequence 1, Appli
5	28	90.3	311	US-09-100-851-1	Sequence 1, Appli
6	28	90.3	311	US-09-265-294-1	Sequence 1, Appli
7	28	90.3	554	US-09-198-452A-1137	Sequence 1137, Ap
8	28	90.3	860	US-08-117-362-3	Sequence 3, Appli
9	28	90.3	860	US-08-486-924-3	Sequence 3, Appli
10	28	90.3	860	US-08-486-929A-3	Sequence 3, Appli
11	27	87.1	177	US-08-469-260A-53	Sequence 53, Appl
12	27	87.1	177	US-08-488-446-53	Sequence 53, Appl
13	27	87.1	177	US-08-467-344A-53	Sequence 53, Appl
14	27	87.1	225	US-09-252-991A-26831	Sequence 26831, A
15	27	87.1	334	US-09-252-991A-26831	Patent No. 5290690
16	27	87.1	335	US-09-252-991A-26831	Patent No. 5290690
17	27	87.1	347	US-08-857-076-100	Sequence 100, App
18	27	87.1	363	US-09-046-086-2	Sequence 2, Appli
19	27	87.1	363	US-09-524-643-2	Sequence 2, Appli
20	27	87.1	556	US-08-687-590-24	Sequence 24, Appli
21	27	87.1	894	US-08-117-362-4	Sequence 4, Appli
22	27	87.1	894	US-08-486-924-4	Sequence 4, Appli
23	27	87.1	894	US-08-486-929A-4	Sequence 4, Appli
24	27	87.1	971	US-09-112-450-2	Sequence 2, Appli
25	27	87.1	971	US-09-419-291A-2	Sequence 2, Appli
26	27	87.1	1054	US-07-596-467-4	Sequence 4, Appli
27	27	87.1	1054	US-07-934-374-4	Sequence 4, Appli

28	27	87.1	1054	1	US-07-783-861C-2	Sequence 2, Appli
29	27	87.1	1430	4	US-09-252-991A-18190	Sequence 18190, A
30	27	87.1	2471	3	US-09-112-450-4	Sequence 4, Appli
31	27	87.1	2471	4	US-09-419-291A-4	Sequence 4, Appli
32	27	87.1	2972	4	US-08-469-260A-387	Sequence 387, App
33	27	87.1	2972	4	US-08-488-446-387	Sequence 387, App
34	27	87.1	2972	4	US-08-467-344A-387	Sequence 387, App
35	26	83.9	39	4	US-09-149-476-644	Sequence 644, App
36	26	83.9	100	4	US-09-732-210-1176	Sequence 1176, Ap
37	26	83.9	215	4	US-09-252-991A-31872	Sequence 31872, A
38	26	83.9	269	4	US-09-443-041A-28	Sequence 28, Appli
39	26	83.9	271	4	US-09-688-019-4	Sequence 4, Appli
40	26	83.9	304	4	US-09-252-991A-31006	Sequence 31006, A
41	26	83.9	433	4	US-09-595-424-2	Sequence 2, Appli
42	26	83.9	493	4	US-09-328-352-5214	Sequence 5214, Ap
43	26	83.9	494	4	US-09-595-424-4	Sequence 4, Appli
44	26	83.9	524	4	US-09-252-991A-18092	Sequence 18092, A
45	26	83.9	569	4	US-09-362-831-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-328-352-4861
; Sequence 4861, Application US/09328352
; Patent No. 6562958

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4861
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4861

Query Match 100.0%; Score 31; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 67 LVHGKL 72

RESULT 2

US-09-328-352-6053
; Sequence 6053, Application US/09328352
; Patent No. 6562958

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6053
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6053

Query Match 93.5%; Score 29; DB 4; Length 684;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

```
Db          513 LVHGKM 518          |||||:
;
; LIBRARY: BRAINOT09
; CLONE: 2150905
US-08-872-784-1

Query Match          90.3%; Score 28; DB 1; Length 311;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy          1 LVHGKL 6
           |||||:
Db          278 LVHGRL 283

RESULT 5
US-09-100-851-1
; Sequence 1, Application US/09100851
; Patent No. 5911984
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PEROXISOMAL THIOESTERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,851
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/872,784
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0293 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT09
; CLONE: 2150905
US-09-100-851-1

Query Match          90.3%; Score 28; DB 2; Length 311;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy          1 LVHGKL 6
           |||||:
Db          278 LVHGRL 283

RESULT 6
US-09-265-294-1
; Sequence 1, Application US/09265294
; Patent No. 6210890
; GENERAL INFORMATION:
```

APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN PEROXISOMAL THIOESTERASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,294
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/100,851
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0293 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT09
CLONE: 2150905
US-09-265-294-1

Query Match 90.3%; Score 28; DB 3; Length 311;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 278 LVHGRL 283

RESULT 7

US-09-198-452A-1137
Sequence 1137, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:

APPLICANT: Grifffais, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 1137
LENGTH: 554
TYPE: PRT

ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1137

Query Match 90.3%; Score 28; DB 4; Length 554;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 77 IIHGKL 82

RESULT 8

US-08-117-362-3
Sequence 3, Application US/08117362
Patent No. 5595872
GENERAL INFORMATION:

APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 860 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-117-362-3

Query Match 90.3%; Score 28; DB 1; Length 860;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 93 LIHGKI 98

RESULT 9

US-08-486-924-3
Sequence 3, Application US/08486924
Patent No. 5789197
GENERAL INFORMATION:

APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 860 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-924-3

Query Match 90.3%; Score 28; DB 1; Length 860;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 93 LIHGKI 98

RESULT 10
US-08-486-929A-3
Sequence 3, Application US/08486929A
Patent No. 6492365
GENERAL INFORMATION:
APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,929A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 860 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-929A-3

Query Match 90.3%; Score 28; DB 4; Length 860;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 93 LIHGKI 98

RESULT 11
US-08-469-260A-53
Sequence 53, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAMI J. PILOT-MATIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIJK
APPLICANT: ISA K. MUSHAWAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-260A-53

Query Match 87.1%; Score 27; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 161 LVHGK 165

RESULT 12
US-08-488-446-53
; Sequence 53, Application US/08488446
; Patent No. 6558896
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHAWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; NUMBER OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; CORRESPONDENCE ADDRESS:
; ADDRESS: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-446-53

Query Match 87.1%; Score 27; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 161 LVHGK 165

RESULT 13
US-08-467-344A-53
; Sequence 53, Application US/08467344A
; Patent No. 6586568

; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHAWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; NUMBER OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; CORRESPONDENCE ADDRESS:
; ADDRESS: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,344A
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,550
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 53:
US-08-467-344A-53

Query Match 87.1%; Score 27; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 161 LVHGK 165

RESULT 14
US-09-252-991A-26831
; Sequence 26831, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26831
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26831

Query Match 87.1%; Score 27; DB 4; Length 225;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 131 LVHGEL 136

RESULT 15

5290690-10
; Patent No. 5290690
; APPLICANT: MRABET, NADIR; LASTERS, IGNACE; STANSSENS, PATRICK
; MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE
; STABILITY OF PROTEINS
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/398,706
; FILING DATE: 25-AUG-1989
; SEQ ID NO: 10:
; LENGTH: 334
5290690-10

Query Match 87.1%; Score 27; DB 6; Length 334;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6
DB 49 VHGL 53

RESULT 16

5290690-9
; Patent No. 5290690
; APPLICANT: MRABET, NADIR; LASTERS, IGNACE; STANSSENS, PATRICK
; MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE
; STABILITY OF PROTEINS
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/398,706
; FILING DATE: 25-AUG-1989
; SEQ ID NO: 9:
; LENGTH: 335
5290690-9

Query Match 87.1%; Score 27; DB 6; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6
DB 50 VHGL 54

RESULT 17

US-08-857-076-100
; Sequence 100, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth

; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-857-076-100

Query Match 87.1%; Score 27; DB 3; Length 347;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 98 LVHGK 102

RESULT 18

US-09-046-086-2
; Sequence 2, Application US/09046086
; Patent No. 6127147
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: No. 6127147el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,086
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 60/043,489
; APPLICATION NUMBER: 60/043,489
; FILING DATE: 10-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM50025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-046-086-2

Query Match 87.1%; Score 27; DB 3; Length 363;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6
Db 300 VHGL 304

RESULT 19
US-09-524-643-2
; Sequence 2, Application US/09524643
; Patent No. 6498234
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; TITLE OF INVENTION: No. 6498234el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/524,643
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/046,086
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GMS0025
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-524-643-2

Query Match 87.1%; Score 27; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6
Db 300 VHGL 304

RESULT 20
US-08-687-590-24
; Sequence 24, Application US/08687590
; Patent No. 6255070
; GENERAL INFORMATION:
; APPLICANT: Willison, Keith Robert
; APPLICANT: Kubota, Hiroshi
; APPLICANT: Ashworth, Alan
; TITLE OF INVENTION: Folding Proteins
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California

; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,590
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00192
; FILING DATE: 31-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9401791.0
; FILING DATE: 31-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9418234.2
; FILING DATE: 09-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 084619-00000000US
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-687-590-24

Query Match 87.1%; Score 27; DB 3; Length 556;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 490 LVHGK 494

RESULT 21
US-08-117-362-4
; Sequence 4, Application US/08117362
; Patent No. 5595872
; GENERAL INFORMATION:
; APPLICANT: Wetterau II, John R.
; APPLICANT: Sharp, Daru Y.
; APPLICANT: Gregg, Richard E.
; TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,362
; FILING DATE: 03-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC21b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901

TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-117-362-4

Query Match 87.1%; Score 27; DB 1; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

RESULT 22
US-08-486-924-4
Sequence 4, Application US/08486924
Patent No. 5789197
GENERAL INFORMATION:
APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELEPHONE: (609) 252-4526
TELECOMMUNICATION INFORMATION:
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-924-4

Query Match 87.1%; Score 27; DB 1; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

US-08-486-924-4
Sequence 4, Application US/08486924
Patent No. 5789197
GENERAL INFORMATION:
APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELEPHONE: (609) 252-4526
TELECOMMUNICATION INFORMATION:
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-924-4

Query Match 87.1%; Score 27; DB 1; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

RESULT 23
US-08-486-929A-4
Sequence 4, Application US/08486929A
Patent No. 6492365
GENERAL INFORMATION:
APPLICANT: Wetterau II, John R.
APPLICANT: Sharp, Daru Y.
APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,929A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELEPHONE: (609) 252-4526
TELECOMMUNICATION INFORMATION:
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-929A-4

Query Match 87.1%; Score 27; DB 4; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

US-08-486-929A-4
Sequence 4, Application US/09112450
Patent No. 6120999
GENERAL INFORMATION:
APPLICANT: Abad, Antonio Jose C.
APPLICANT: Choi, Gil
APPLICANT: Calderone, Richard A.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: PF393
CURRENT APPLICATION NUMBER: US/09/112,450
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/052,273
EARLIER FILING DATE: 1997-07-10
EARLIER APPLICATION NUMBER: 60/074,308
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2

Query Match 87.1%; Score 27; DB 4; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

US-09-112-450-2
Sequence 2, Application US/09112450
Patent No. 6120999
GENERAL INFORMATION:
APPLICANT: Abad, Antonio Jose C.
APPLICANT: Choi, Gil
APPLICANT: Calderone, Richard A.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: PF393
CURRENT APPLICATION NUMBER: US/09/112,450
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/052,273
EARLIER FILING DATE: 1997-07-10
EARLIER APPLICATION NUMBER: 60/074,308
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2

Query Match 87.1%; Score 27; DB 4; Length 894;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 127 LIHGKV 132

; LENGTH: 971
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-112-450-2

Query Match 87.1%; Score 27; DB 3; Length 971;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|||||
Db 284 LVHGK 288

RESULT 25

US-09-419-291A-2
; Sequence 2, Application US/09419291A
; Patent No. 6416989
; GENERAL INFORMATION:

; APPLICANT: Abad et al.
; TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
; FILE REFERENCE: PB393D1
; CURRENT APPLICATION NUMBER: US/09/419,291A
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: US 09/112,450
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: US 60/074,308
; PRIOR FILING DATE: 1998-02-11
; PRIOR APPLICATION NUMBER: US 60/052,273
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-419-291A-2

Query Match 87.1%; Score 27; DB 4; Length 971;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|||||
Db 284 LVHGK 288

RESULT 26

US-07-596-467-4
; Sequence 4, Application US/07596467
; Patent No. 5306862
; GENERAL INFORMATION:

; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.
; TITLE OF INVENTION: Method and Composition for Increasing
; Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/596,467

; FILING DATE: 19901012
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1054 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-596-467-4

Query Match 87.1%; Score 27; DB 1; Length 1054;
Best Local Similarity 66.7%; Pred. No. 8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 625 VIHGL 630

RESULT 27

US-07-934-374-4
; Sequence 4, Application US/07934374
; Patent No. 5349126
; GENERAL INFORMATION:

; APPLICANT: Chappell, J.
; APPLICANT: Saunders, Court A.
; APPLICANT: Cuellar, R.
; APPLICANT: Wolf, Fred R.
; TITLE OF INVENTION: Process and Composition for Increasing
; Sterol Accumulation in Higher Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker & Milnamow
; STREET: 180 N. Stetson St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/934,374
; FILING DATE: 19920814
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 616-5400
; TELEFAX: 312 616-5460
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1054 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-934-374-4

Query Match 87.1%; Score 27; DB 1; Length 1054;
Best Local Similarity 66.7%; Pred. No. 8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 625 VIHGL 630

```
RESULT 28
US-07-783-861C-2
; Sequence 2, Application US/07783861C
; Patent No. 5460949
; GENERAL INFORMATION:
; APPLICANT: Saunders, Court A.
; APPLICANT: Wolf, Fred R.
; APPLICANT: Mukharji, Indrani
; TITLE OF INVENTION: A Method and Composition for Increasing
; TITLE OF INVENTION: the Accumulation of Squalene and Specific Sterols in
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept.
; STREET: 200 East Randolph St.
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07783.861C
; FILING DATE: 19911028
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,380
; FILING DATE: 15-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5460949vall B.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 856-7180
; TELEFAX: 312 856-4972
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1054 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-861C-2

Query Match      87.1%; Score 27; DB 1; Length 1054;
Best Local Similarity 66.7%; Pred. No. 8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LVHGKL 6
Db      625 VHGKL 630

RESULT 29
US-09-252-991A-18190
; Sequence 18190, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18190
; LENGTH: 1430
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18190

Query Match      87.1%; Score 27; DB 4; Length 1430;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 VHGKL 6
Db      869 VHGKL 873

RESULT 30
US-09-112-450-4
; Sequence 4, Application US/09112450
; Patent No. 6120999
; GENERAL INFORMATION:
; APPLICANT: Abad, Antonio Jose C.
; APPLICANT: Choi, Gil
; APPLICANT: Calderone, Richard A.
; TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
; FILE REFERENCE: PF393
; CURRENT APPLICATION NUMBER: US/09/112.450
; CURRENT FILING DATE: 1998-07-09
; EARLIER APPLICATION NUMBER: 60/052,273
; EARLIER FILING DATE: 1997-07-10
; EARLIER APPLICATION NUMBER: 60/074,308
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2471
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-112-450-4

Query Match      87.1%; Score 27; DB 3; Length 2471;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LVHGK 5
Db      1784 LVHGK 1788

RESULT 31
US-09-419-291A-4
; Sequence 4, Application US/09419291A
; Patent No. 6416989
; GENERAL INFORMATION:
; APPLICANT: Abad et al.
; TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
; FILE REFERENCE: PB393D1
; CURRENT APPLICATION NUMBER: US/09/419,291A
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: US 09/112,450
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: US 60/074,308
; PRIOR FILING DATE: 1998-02-11
; PRIOR APPLICATION NUMBER: US 60/052,273
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2471
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-419-291A-4

Query Match      87.1%; Score 27; DB 4; Length 2471;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LVHGK 5
11111
Db 1784 LVHGK 1788

RESULT 32

US-08-469-260A-387
; Sequence 387, Application US/08469260A

; Patent No. 6451578

; GENERAL INFORMATION:

; APPLICANT: JOHN N. SIMONS

; APPLICANT: TAMI J. PILOT-MATIAS

; APPLICANT: GEORGE J. DAWSON

; APPLICANT: GEORGE G. SCHLAUDER

; APPLICANT: SURESH M. DESAI

; APPLICANT: THOMAS P. LEARY

; APPLICANT: ANTHONY SCOTT MUERHOFF

; APPLICANT: JAMES C. ERKER

; APPLICANT: SHERI L. BUIJK

; APPLICANT: ISA K. MUSHAWAR

; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS

; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE

; NUMBER OF SEQUENCES: 716

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: 100 ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,260A

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/424,550

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: POREMSKI, PRISCILLA E.

; REGISTRATION NUMBER: 33,207

; REFERENCE/DOCKET NUMBER: 5527.PC.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-6365

; TELEFAX: 708-938-2623

; INFORMATION FOR SEQ ID NO: 387:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2972 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-469-260A-387

Query Match 87.1%; Score 27; DB 4; Length 2972;

Best Local Similarity 100.0%; Pred. No. 2.3e+03;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5

11111

Db 949 LVHGK 953

RESULT 33

US-08-488-446-387

; Sequence 387, Application US/08488446

; Patent No. 6558898

; GENERAL INFORMATION:

; APPLICANT: JOHN N. SIMONS

; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHAWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 387:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-446-387

Query Match 87.1%; Score 27; DB 4; Length 2972;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5

11111

Db 949 LVHGK 953

RESULT 34

US-08-467-344A-387

; Sequence 387, Application US/08467344A

; Patent No. 6586568

; GENERAL INFORMATION:

; APPLICANT: JOHN N. SIMONS

; APPLICANT: TAMI J. PILOT-MATIAS

; APPLICANT: GEORGE J. DAWSON

; APPLICANT: GEORGE G. SCHLAUDER

; APPLICANT: SURESH M. DESAI

; APPLICANT: THOMAS P. LEARY

; APPLICANT: ANTHONY SCOTT MUERHOFF

; APPLICANT: JAMES C. ERKER

; APPLICANT: SHERI L. BUIJK

; APPLICANT: ISA K. MUSHAWAR

; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS

; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE

; NUMBER OF SEQUENCES: 716

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,344A
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,550
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 387:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 387:
US-08-467-344A-387

Query Match 87.1%; Score 27; DB 4; Length 2972;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 949 LVHGK 953

RESULT 35
US-09-149-476-644
; Sequence 644, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-03-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 83.9%; Score 26; DB 4; Length 39;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 24 IVHGKV 29

RESULT 36

US-09-732-210-1176
Sequence 1176, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunkers, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Mittanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 1176
LENGTH: 100
TYPE: PRT
ORGANISM: Spinacia oleracea
US-09-732-210-1176

Query Match 83.9%; Score 26; DB 4; Length 100;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
Db 43 IHGKL 47

RESULT 37

US-09-252-991A-31872

```
; Sequence 31872, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31872
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31872

Query Match      83.9%; Score 26; DB 4; Length 215;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 7 VVHGKI 12

RESULT 38
US-09-443-041A-28
; Sequence 28, Application US/09443041A
; Patent No. 6465717
; GENERAL INFORMATION:
; APPLICANT: Famodu, Onclayo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Shen, Jennie
; TITLE OF INVENTION: Sterol Metabolism Enzymes
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/443,041A
; CURRENT FILING DATE: 1999-11-18
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: 60/109,283
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Glycine max
US-09-443-041A-28

Query Match      83.9%; Score 26; DB 4; Length 269;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGKL 6
Db 220 IHGKL 224

RESULT 39
US-09-688-019-4
; Sequence 4, Application US/09688019
; Patent No. 6566512
; GENERAL INFORMATION:
; APPLICANT: Sturley, Stephen L.
; APPLICANT: Tinkelenberg, Arthur H.
; TITLE OF INVENTION: ARVI, A PROTEIN INVOLVED IN STEROL UPTAKE AND STEROL HOMEOSTASIS
; TITLE OF INVENTION: BUDDING YEAST, S. CEREVISIAE, AND A FUNCTIONAL HUMAN ARVI
; FILE REFERENCE: 0575/58072
; CURRENT APPLICATION NUMBER: US/09/688,019
; CURRENT FILING DATE: 2002-04-15
```

```
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Human
US-09-688-019-4

Query Match      83.9%; Score 26; DB 4; Length 271;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGKL 6
Db 99 IHGKL 103

RESULT 40
US-09-252-991A-31006
; Sequence 31006, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31006
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31006

Query Match      83.9%; Score 26; DB 4; Length 304;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 38 LVHGSL 43

RESULT 41
US-09-595-424-2
; Sequence 2, Application US/09595424
; Patent No. 6448025
; GENERAL INFORMATION:
; APPLICANT: Vaisberg, Eugenii
; APPLICANT: Wood, Kenneth
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: No. 6448025el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1041
; CURRENT APPLICATION NUMBER: US/09/595,424
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Human
US-09-595-424-2
```


Query Match 83.9%; Score 26; DB 4; Length 433;
Best Local Similarity 80.0%; Pred. No. 5.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 98 LIHGK 102

RESULT 42

US-09-328-352-5214
; Sequence 5214, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5214
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5214

Query Match 83.9%; Score 26; DB 4; Length 493;
Best Local Similarity 80.0%; Pred. No. 6.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
Db 241 IHGKL 245

RESULT 43

US-09-595-424-4
; Sequence 4, Application US/09595424
; Patent No. 6448025
; GENERAL INFORMATION:
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: No. 6448025el motor proteins and methods for
; FILE REFERENCE: 1041
; CURRENT APPLICATION NUMBER: US/09/595,424
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Human
US-09-595-424-4

Query Match 83.9%; Score 26; DB 4; Length 494;
Best Local Similarity 80.0%; Pred. No. 6.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 98 LIHGK 102

RESULT 44

US-09-252-991A-18092
; Sequence 18092, Application US/09252991A

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18092
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18092

Query Match 83.9%; Score 26; DB 4; Length 524;
Best Local Similarity 66.7%; Pred. No. 6.5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 333 ILHGKL 338

RESULT 45

US-09-362-831-9
; Sequence 9, Application US/09362831
; Patent No. 6306400
; GENERAL INFORMATION:
; APPLICANT: BUBLOT et al.
; TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING, AS VECTOR, THE
; FILE REFERENCE: 454313-2520
; CURRENT APPLICATION NUMBER: US/09/362,831
; CURRENT FILING DATE: 1999-07-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 9
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Newcastle Disease Virus HN
US-09-362-831-9

Query Match 83.9%; Score 26; DB 4; Length 569;
Best Local Similarity 66.7%; Pred. No. 7.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 270 MVHGRL 275

Search completed: October 14, 2003, 13:11:58
Job time : 16.8182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:12:07 ; Search time 24.5455 Seconds
(without alignments)
39.387 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29
Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	60	11	US-09-764-891-3073
2	30	96.8	389	9	US-09-815-242-10058
3	30	96.8	389	9	US-09-815-242-10058
4	28	90.3	92	12	US-10-207-780-97
5	28	90.3	311	9	US-09-766-366-1
6	28	90.3	406	15	US-10-156-761-14349
7	28	90.3	729	9	US-09-741-149-4
8	28	90.3	729	12	US-10-385-614-4
9	28	90.3	730	9	US-09-741-149-2
10	28	90.3	730	9	US-09-795-693-5
11	28	90.3	730	12	US-10-385-614-2
12	28	90.3	730	15	US-10-156-239-5
13	28	90.3	730	15	US-10-198-485-5
14	28	90.3	759	15	US-10-128-714-3162
15	28	90.3	869	15	US-10-128-714-8162

```

16 87.1 46 9 US-09-864-761-35096
17 87.1 69 15 US-10-198-070-32
18 87.1 177 8 US-08-424-5508-53
19 87.1 184 11 US-09-873-546-15
20 87.1 195 15 US-10-156-761-12656
21 87.1 196 12 US-09-738-269-25
22 87.1 196 14 US-10-023-437-25
23 87.1 226 9 US-09-925-301-1023
24 87.1 294 9 US-09-815-242-5790
25 87.1 299 9 US-09-815-242-10697
26 87.1 321 15 US-10-156-761-8413
27 87.1 347 9 US-09-844-353A-100
28 87.1 363 9 US-09-815-242-12801
29 87.1 475 9 US-09-815-242-11259
30 87.1 598 9 US-09-815-242-11723
31 87.1 643 10 US-09-736-968A-60
32 87.1 649 12 US-09-738-269-27
33 87.1 649 14 US-10-023-437-27
34 87.1 675 11 US-09-978-244A-18
35 87.1 717 15 US-10-101-464A-810
36 87.1 732 15 US-10-156-761-10109
37 87.1 971 14 US-10-116-048-2
38 87.1 999 15 US-10-157-031-26
39 87.1 1236 12 US-10-023-649-4
40 87.1 1236 12 US-10-023-649-6
41 87.1 1874 12 US-10-238-075-961
42 87.1 2045 9 US-09-736-969A-95
43 87.1 2045 10 US-09-736-960-92
44 87.1 2045 10 US-09-736-968A-109
45 87.1 2047 10 US-09-736-968A-2

```

ALIGNMENTS

RESULT 1

```

US-09-764-891-3073
; Sequence 3073, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3073
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3073

```

Query Match 100.0%; Score 31; DB 11; Length 60;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 LVHGKL 6
    |||||
Db 44 LVHGKL 49

RESULT 2
US-09-815-242-10058
; Sequence 10058, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10058
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10058

Query Match 96.8%; Score 30; DB 9; Length 389;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
    |||||
Db 65 LVHGKL 70

RESULT 3
US-09-815-242-13894
; Sequence 13894, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

Query Match 96.8%; Score 30; DB 9; Length 389;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
    |||||
Db 65 LVHGKL 70

RESULT 4
US-10-207-780-97
; Sequence 97, Application US/10207780
; Publication No. US20030144210A1
; GENERAL INFORMATION:
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Walker, Craig S.
; APPLICANT: Watkins, Maren
; APPLICANT: Jones, Robert M.
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; TITLE OF INVENTION: Gamma-Carboxyglutamate Containing Conopeptides
; FILE REFERENCE: Gla-Conopeptides
; CURRENT APPLICATION NUMBER: US/10/207,780
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: US/09/658,603
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/219,673
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/153,034
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 97
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Conus cinereus gubba
US-10-207-780-97

Query Match 90.3%; Score 28; DB 12; Length 92;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LVHGKL 6
    |||||
Db 87 LVHGKI 92

RESULT 5
US-09-766-366-1
```

; Sequence 1, Application US/09766366
; Patent No. US20010024786A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Shah, Furvi
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN PEROXISOMAL THIOESTERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/766,366
; FILING DATE: 18-Jan-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/872,784
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0293 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT09
; CLONE: 2150905
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-766-366-1

Query Match 90.3%; Score 28; DB 9; Length 311;
Best Local Similarity 83.3%; Pred. No. 4.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 278 LVHGRL 283

RESULT 6
US-10-156-761-14349
; Sequence 14349, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBU, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14349
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14349

Query Match 90.3%; Score 28; DB 15; Length 406;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 355 LVHGRL 360

RESULT 7
US-09-741-149-4
; Sequence 4, Application US/09741149
; Patent No. US20020031800A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000780
; CURRENT APPLICATION NUMBER: US/09/741,149
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-741-149-4

Query Match 90.3%; Score 28; DB 9; Length 729;
Best Local Similarity 66.7%; Pred. No. 9.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 676 LVHGKI 681

RESULT 8
US-10-385-614-4
; Sequence 4, Application US/10385614
; Publication No. US20030157649A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000780
; CURRENT APPLICATION NUMBER: US/10/385,614
; CURRENT FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-385-614-4

Query Match 90.3%; Score 28; DB 12; Length 729;
Best Local Similarity 66.7%; Pred. No. 9.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|:
Db 676 LVHGKI 681

```

RESULT 9
US-09-741-149-2
; Sequence 2, Application US/09741149
; Patent No. US20020031800A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000780
; CURRENT APPLICATION NUMBER: US/09/741,149
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 730
; TYPE: PRT
; ORGANISM: Human
US-09-741-149-2

Query Match          90.3%;   Score 28;   DB 9;   Length 730;
Best Local Similarity 66.7%;   Pred. No. 9.7e+02;
Matches 4;   Conservative 2;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1 LVHGKL 6
DB      676 LIHGKI 681

RESULT 10
US-09-795-693-5
; Sequence 5, Application US/09795693
; Patent No. US20020068710A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 20685, 579, 17114, 23821, 33894, and
; TITLE OF INVENTION: 32613, No. US20020068710A1el Human Transporters
; FILE REFERENCE: 35800/209292
; CURRENT APPLICATION NUMBER: US/09/795,693
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 730
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-693-5

Query Match          90.3%;   Score 28;   DB 9;   Length 730;
Best Local Similarity 66.7%;   Pred. No. 9.7e+02;
Matches 4;   Conservative 2;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1 LVHGKL 6
DB      676 LIHGKI 681

RESULT 11
US-10-385-614-2
; Sequence 2, Application US/10385614
; Publication No. US20030157649A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000780
; CURRENT APPLICATION NUMBER: US/10/385,614
; CURRENT FILING DATE: 2003-03-12

```

; CURRENT APPLICATION NUMBER: US/10/199,485
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 09/795,693
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,906
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 730
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-485-5

Query Match 90.3%; Score 28; DB 15; Length 730;
Best Local Similarity 66.7%; Pred. No. 9.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 676 LIRGKI 681

RESULT 14

US-10-128-714-3162
; Sequence 3162, Application US/10128714
; Publication No. US20030119013A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wenqi

; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos

; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lemieux, Sebastien M

; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; FILE REFERENCE: 10182-018-999

; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3162

; LENGTH: 759

; TYPE: PRT

; ORGANISM: Aspergillus fumigatus

US-10-128-714-3162

Query Match 90.3%; Score 28; DB 15; Length 759;
Best Local Similarity 83.3%; Pred. No. 1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 43 VVHGKL 48

RESULT 15

US-10-128-714-8162

; Sequence 8162, Application US/10128714

; Publication No. US20030119013A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wenqi

; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8162
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8162

Query Match 90.3%; Score 28; DB 15; Length 869;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 175 VVHGKL 180

RESULT 16

US-09-864-761-35096

; Sequence 35096, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 35096
;; LENGTH: 46
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL035691.17
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
;; OTHER INFORMATION: SWISSPROT HIT: O28392, EVALUE 6.00e+00
;; OTHER INFORMATION: EST_HUMAN HIT: AA088605.1, EVALUE 4.00e-01
US-09-864-761-35096

Query Match 87.1%; Score 27; DB 9; Length 46;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 12 VIHGL 17

RESULT 17
US-10-198-070-32
; Sequence 32, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GEMMELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 59003-000008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 32
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-070-32

Query Match 87.1%; Score 27; DB 15; Length 69;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 32 LAHGKL 37

RESULT 18

US-08-424-550B-53
; Sequence 53, Application US/08424550B
; Publication No. US20020119447A1
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSAHWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-550B-53

Query Match 87.1%; Score 27; DB 8; Length 177;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGK 5
Db 161 LVHGK 165

RESULT 19

US-09-873-546-15
; Sequence 15, Application US/09873546
; Publication No. US20030059771A1
; GENERAL INFORMATION:
; APPLICANT: Clark, Geoff
; APPLICANT: Ellis, Chad

; APPLICANT: Vos, Michelle
; TITLE OF INVENTION: Rig: No. US20030059771a1el Ras-Related Gene
; FILE REFERENCE: NIH-05080
; CURRENT APPLICATION NUMBER: US/09/873,546
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-873-546-15

Query Match 87.1%; Score 27; DB 11; Length 184;
Best Local Similarity 66.7%; Pred. No. 4.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 98 VIHGL 103

RESULT 20
US-10-156-761-12656
; Sequence 12656, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12656
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12656

Query Match 87.1%; Score 27; DB 15; Length 195;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 76 LVHGL 81

RESULT 21
US-09-738-269-25
; Sequence 25, Application US/09738269
; Publication No. US20030185848A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: STEMKE-HALE, KATHERINE
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: KALTENBOECK, BERNHARD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF
; FILE REFERENCE: UTSD:659
; CURRENT APPLICATION NUMBER: US/09/738,269
; CURRENT FILING DATE: 2000-12-18

; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Chlamydia psittaci
US-09-738-269-25

Query Match 87.1%; Score 27; DB 12; Length 196;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 116 LVHGK 120

RESULT 22
US-10-023-437-25
; Sequence 25, Application US/10023437
; Publication No. US20020183272A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: STEMKE-HALE, KATHERINE
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: KALTENBOECK, BERNHARD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING NUCLEIC AC
; TITLE OF INVENTION: AND/OR POLYPEPTIDE SEQUENCES OF CHLAMYDIA
; FILE REFERENCE: UTSD:736US
; CURRENT APPLICATION NUMBER: US/10/023,437
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/225,839
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Chlamydia psittaci
US-10-023-437-25

Query Match 87.1%; Score 27; DB 14; Length 196;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 116 LVHGK 120

RESULT 23
US-09-925-301-1023
; Sequence 1023, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1023
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (33)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1023

Query Match 87.1%; Score 27; DB 9; Length 226;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
| | | | : |
Db 15 LVHGEL 20

RESULT 24

US-09-815-242-5790
; Sequence 5790, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5790

; LENGTH: 294

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5790

Query Match 87.1%; Score 27; DB 9; Length 294;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGL 6
| | | | |
Db 237 VHGL 241

RESULT 25

US-09-815-242-10697

; Sequence 10697, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 10697

; LENGTH: 299

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-10697

Query Match 87.1%; Score 27; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGK 5
| | | | |
Db 168 LVHGK 172

RESULT 26

US-10-156-761-8413

; Sequence 8413, Application US/10156761

; Publication No. US20030119018A1

GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 8413

; LENGTH: 321

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-8413

Query Match 87.1%; Score 27; DB 15; Length 321;
Best Local Similarity 66.7%; Pred. No. 7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|
Db 265 LHHGL 270

RESULT 27

US-09-844-353A-100
; Sequence 100, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-844-353A-100

Query Match 87.1%; Score 27; DB 9; Length 347;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|:|:|
Db 98 LVHGK 102

RESULT 28

US-09-815-242-112801
; Sequence 12801, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12801
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12801

Query Match 87.1%; Score 27; DB 9; Length 363;
Best Local Similarity 100.0%; Pred. No. 7.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
|:|:|
Db 300 VHGKL 304

RESULT 29

US-09-815-242-11259
; Sequence 11259, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11259
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11259

Query Match 87.1%; Score 27; DB 9; Length 475;
Best Local Similarity 83.3%; Pred. No. 1e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|:|
Db 4 LVHGKL 9

RESULT 30

US-09-815-242-11723
; Sequence 11723, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11723
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11723

Query Match 87.1%; Score 27; DB 9; Length 598;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGLK 6
Db 113 VHGLK 117

RESULT 31
US-09-736-968A-60
; Sequence 60, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
; FILE REFERENCE: 020054-000611US
; CURRENT APPLICATION NUMBER: US/09/736,968A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11

; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: preliminary human CLASP-7 cDNA sequence
US-09-736-968A-60

Query Match 87.1%; Score 27; DB 10; Length 643;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGLK 6
Db 322 VHGLK 326

RESULT 32
US-09-738-269-27
; Sequence 27, Application US/09738269
; Publication No. US20030185848A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: KALTENBOECK, BERNHARD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF
; FILE REFERENCE: UTSD-659
; CURRENT APPLICATION NUMBER: US/09/738,269
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Chlamydia psittaci
US-09-738-269-27

Query Match 87.1%; Score 27; DB 12; Length 649;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHCK 5
Db 547 LVHCK 551

RESULT 33
US-10-023-437-27
; Sequence 27, Application US/10023437
; Publication No. US20020183272A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: KALTENBOECK, BERNHARD

;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING NUCLEIC ACID
;; FILE OF INVENTION: AND/OR POLYPEPTIDE SEQUENCES OF CHLAMYDIA

;; FILE REFERENCE: UTSD:736US
;; CURRENT APPLICATION NUMBER: US/10/023,437

;; CURRENT FILING DATE: 2001-12-17

;; PRIOR APPLICATION NUMBER: 60/225,839

;; PRIOR FILING DATE: 2000-12-15

;; NUMBER OF SEQ ID NOS: 69

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 27

;; LENGTH: 649

;; TYPE: PRT

;; ORGANISM: Chlamydia psittaci

US-10-023-437-27

Query Match 87.1%; Score 27; DB 14; Length 649;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5

|||||

Db 547 LVHGK 551

RESULT 34

US-09-978-244A-18

;; Sequence 18, Application US/09978244A

;; Publication No. US20030103992A1

;; GENERAL INFORMATION:

;; APPLICANT: Lu, Peter S

;; APPLICANT: Garman, Jonathan D.

;; APPLICANT: Candia III, Albert F.

;; APPLICANT: Arbor Vita Corporation

;; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS

;; FILE REFERENCE: 020554-000161US

;; CURRENT APPLICATION NUMBER: US/09/978,244A

;; CURRENT FILING DATE: 2001-10-15

;; PRIOR APPLICATION NUMBER: US 60/310,028

;; PRIOR FILING DATE: 2001-08-03

;; PRIOR APPLICATION NUMBER: US 09/737,246

;; PRIOR FILING DATE: 2000-12-13

;; PRIOR APPLICATION NUMBER: US 09/736,969

;; PRIOR FILING DATE: 2000-12-13

;; PRIOR APPLICATION NUMBER: US 09/736,960

;; PRIOR FILING DATE: 2000-12-13

;; PRIOR APPLICATION NUMBER: US 09/736,968

;; PRIOR FILING DATE: 2000-12-13

;; PRIOR APPLICATION NUMBER: US 60/240,545

;; PRIOR FILING DATE: 2000-10-13

;; PRIOR APPLICATION NUMBER: US 60/240,508

;; PRIOR FILING DATE: 2000-10-13

;; PRIOR APPLICATION NUMBER: US 60/240,503

;; PRIOR FILING DATE: 2000-10-13

;; PRIOR APPLICATION NUMBER: US 60/240,539

;; PRIOR FILING DATE: 2000-10-13

;; PRIOR APPLICATION NUMBER: US 60/240,543

;; PRIOR FILING DATE: 2000-10-13

;; Remaining Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 106

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 18

;; LENGTH: 675

;; TYPE: PRT

;; ORGANISM: Mus sp.

;; FEATURE:

;; OTHER INFORMATION: CLASP-7 Fragment 3

US-09-978-244A-18

Query Match 87.1%; Score 27; DB 11; Length 675;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGLK 6

Db 354 VHGLK 358

|||||

RESULT 35

US-10-101-464A-810

;; Sequence 810, Application US/10101464A

;; Publication No. US20030046728A1

;; GENERAL INFORMATION:

;; APPLICANT: Strabala, Timothy

;; APPLICANT: Nieuwenhuizen, Nicolaas

;; APPLICANT: Higgins, Colleen M.

;; TITLE OF INVENTION: Compositions Isolated from Plant Cells

;; FILE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling

;; FILE REFERENCE: 11000.1020c2

;; CURRENT APPLICATION NUMBER: US/10/101,464A

;; CURRENT FILING DATE: 2002-03-18

;; PRIOR APPLICATION NUMBER: 09/704,302

;; PRIOR FILING DATE: 2000-11-01

;; PRIOR APPLICATION NUMBER: 09/228,986

;; PRIOR FILING DATE: 1999-01-12

;; PRIOR APPLICATION NUMBER: 60/162,866

;; PRIOR FILING DATE: 1999-11-01

;; PRIOR APPLICATION NUMBER: PCT/US00/00724

;; PRIOR FILING DATE: 2000-01-11

;; NUMBER OF SEQ ID NOS: 989

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 810

;; LENGTH: 717

;; TYPE: PRT

;; ORGANISM: Eucalyptus grandis

US-10-101-464A-810

Query Match

87.1%; Score 27; DB 15; Length 717;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGLK 6

|||||

Db 533 VHGLK 537

RESULT 36

US-10-156-761-10109

;; Sequence 10109, Application US/10156761

;; Publication No. US20030119018A1

;; GENERAL INFORMATION:

;; APPLICANT: OMURA, SATOSHI

;; APPLICANT: IKEDA, HARUO

;; APPLICANT: ISHIKAWA, JUN

;; APPLICANT: HORIKAWA, HIROSHI

;; APPLICANT: SHIBA, TADAYOSHI

;; APPLICANT: SAKAKI, YOSHIYUKI

;; APPLICANT: HATTORI, MASAHIRA

;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

;; FILE REFERENCE: 249-262

;; CURRENT APPLICATION NUMBER: US/10/156,761

;; CURRENT FILING DATE: 2002-05-29

;; PRIOR APPLICATION NUMBER: JP 2001-204089

;; PRIOR FILING DATE: 2001-05-30

;; PRIOR APPLICATION NUMBER: JP 2001-272697

;; PRIOR FILING DATE: 2001-08-02

;; NUMBER OF SEQ ID NOS: 15109

;; SEQ ID NO 10109

;; LENGTH: 732

;; TYPE: PRT

;; ORGANISM: Streptomyces avermitilis

US-10-156-761-10109

Query Match

87.1%; Score 27; DB 15; Length 732;

Best Local Similarity 83.3%; Pred. No. 1.6e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
 ||||:1
Db 372 LVHGQL 377

RESULT 37

US-10-116-048-2
; Sequence 2, Application US/10116048
; Publication No. US20020146738A1
; GENERAL INFORMATION:
; APPLICANT: Abad et al.
; TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
; FILE REFERENCE: PB393D2
; CURRENT APPLICATION NUMBER: US/10/116.048
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US 09/419,291
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: US 09/112,450
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: US 60/074,308
; PRIOR FILING DATE: 1998-02-11
; PRIOR APPLICATION NUMBER: US 60/052,273
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-116-048-2

Query Match 87.1%; Score 27; DB 14; Length 971;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
 |||||
Db 284 LVHGK 288

RESULT 38

US-10-157-031-26
; Sequence 26, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-26

Query Match 87.1%; Score 27; DB 15; Length 999;
Best Local Similarity 83.3%; Pred. No. 2.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
 ||||:1
Db 20 LVHGEL 25

RESULT 39

US-10-023-649-4

; Sequence 4, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
; APPLICANT: The Minister of National Defence, Government of Canada
; APPLICANT: Nagata, Leslie P
; APPLICANT: Wong, Jonathan P
; TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalit
; FILE REFERENCE: NEL-001
; CURRENT APPLICATION NUMBER: US/10/023,649
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: 60/256,948
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
US-10-023-649-4

Query Match 87.1%; Score 27; DB 12; Length 1236;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
 |||||
Db 464 VHGKL 468

RESULT 40

US-10-023-649-6
; Sequence 6, Application US/10023649
; Publication No. US20030143201A1
; GENERAL INFORMATION:
; APPLICANT: The Minister of National Defence, Government of Canada
; APPLICANT: Nagata, Leslie P
; APPLICANT: Wong, Jonathan P
; TITLE OF INVENTION: No. US20030143201A1el DNA-Based Vaccine Against the Encephalit
; FILE REFERENCE: NEL-001
; CURRENT APPLICATION NUMBER: US/10/023,649
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: 60/256,948
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Western equine encephalomyelitis virus - strain 71V-1658
US-10-023-649-6

Query Match 87.1%; Score 27; DB 12; Length 1236;
Best Local Similarity 100.0%; Pred. No. 2.6e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
 |||||
Db 464 VHGKL 468

RESULT 41

US-10-238-075-961
; Sequence 961, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I. N. S. E. R. M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are iso
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 961
; LENGTH: 1874
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-961

Query Match 87.1%; Score 27; DB 12; Length 1874;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 601 LVHGK 605

RESULT 42
US-09-736-969A-95
; Sequence 95, Application US/09736969A
; Patent No. US20020068302A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-4 Transmembrane Protein
; FILE REFERENCE: 020054-000411US
; CURRENT APPLICATION NUMBER: US/09/736,969A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 95
; LENGTH: 2045
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-7
US-09-736-969A-95

Query Match 87.1%; Score 27; DB 9; Length 2045;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
Db 1724 VHGKL 1728

RESULT 43
US-09-736-960-92
; Sequence 92, Application US/09736960
; Patent No. US20020102267A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-5 Transmembrane Protein
; FILE REFERENCE: 020054-000511US
; CURRENT APPLICATION NUMBER: US/09/736,960
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 92
; LENGTH: 2045
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-7
US-09-736-960-92

Query Match 87.1%; Score 27; DB 10; Length 2045;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
Db 1724 VHGKL 1728

RESULT 44
US-09-736-968A-109
; Sequence 109, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
US-09-736-968A-109

; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
; FILE REFERENCE: 020054-000611US
; CURRENT APPLICATION NUMBER: US/09/736,968A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 2045
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-7
US-09-736-968A-109

Query Match 87.1%; Score 27; DB 10; Length 2045;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGL 6
Db 1724 VHGL 1728

RESULT 45
US-09-736-968A-2
; Sequence 2, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
; FILE REFERENCE: 020054-000611US
; CURRENT APPLICATION NUMBER: US/09/736,968A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498

; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length human CLASP-7
US-09-736-968A-2

Query Match 87.1%; Score 27; DB 10; Length 2047;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGL 6
Db 1726 VHGL 1730

Search completed: October 14, 2003, 13:16:18
Job time : 29.5455 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 14.1818 Seconds
(without alignments)
40.687 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29

Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No: is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	553	2 AH1830	serine/threonine k
2	31	100.0	767	2 T07693	hypothetical prote
3	30	96.8	389	2 E64760	citrate (sl)-synth
4	30	96.8	389	2 B90677	probable citrate s
5	30	96.8	389	2 E85527	hypothetical prote
6	30	96.8	389	2 AH0547	methylocitrate synt
7	29	93.5	107	2 B42529	B-ORF-A protein -
8	29	93.5	143	2 T04091	hypothetical prote
9	29	93.5	265	2 B84063	hypothetical prote
10	29	93.5	393	2 B82185	glycerol-3-phospha
11	29	93.5	980	2 T30089	probable zinc prot
12	29	93.5	1131	2 AD2005	RNA polymerase bet
13	28	90.3	178	2 A86370	hypothetical prote
14	28	90.3	188	2 D97260	probable membrane
15	28	90.3	254	2 F84061	hypothetical prote
16	28	90.3	317	2 T01758	hypothetical prote
17	28	90.3	319	1 JC5644	acyl-CoA thioleste
18	28	90.3	383	2 H64396	hypothetical prote
19	28	90.3	484	2 S23817	hypothetical prote
20	28	90.3	554	2 C72002	exodeoxyribonuclea
21	28	90.3	554	2 C86623	exodeoxyribonuclea
22	28	90.3	583	2 T19839	hypothetical prote
23	28	90.3	682	1 H63879	ATP-dependent DNA
24	28	90.3	682	2 AC1301	ATP-dependent DNA
25	28	90.3	682	2 AC1673	ATP-dependent DNA
26	28	90.3	686	2 F89895	ATP-dependent DNA
27	28	90.3	728	2 I63413	sodium-dependent n
28	28	90.3	730	2 I52632	sodium-dependent n
29	28	90.3	764	2 S48521	AKR1 protein - yea

ALIGNMENTS

RESULT 1

AH1830

serine/threonine kinase [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AH1830

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AH1830

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-553 <XUR>

A;Cross-references: GB:BA000019; PIDN:BA077716.1; PID:g17135170; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:
A;Gene: all0192

Query Match 100.0%; Score 31; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 155 LVHGKL 160

RESULT 2

T07693

hypothetical protein F17N18.20 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 22-Oct-1999

C;Accession: T07693

R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Sch

submitted to the Protein Sequence Database, May 1999

A;Reference number: Z16098

A;Accession: T07693

A;Molecule type: DNA

A;Residues: 1-767 <BEV>

A;Cross-references: EMBL:AL049751; GSPDB:GN00062; ATSP:F17N18.20

A;Experimental source: cultivar Columbia; BAC clone F17N18

C;Genetics:

A;Gene: ATSP:F17N18.20

A;Map position: 4

Query Match 100.0%; Score 31; DB 2; Length 767;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 196 LVHGKL 201

RESULT 3
E64760
citrate (si)-synthase (EC 4.1.3.7) - Escherichia coli (strain K-12)
N:Alternate names: citrate synthase II; citrate-condensing enzyme; citrogenase; oxaloacetate synthase; Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E64760; S33286
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64760
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-389 <BLAT>
A:Cross-references: GB:AF000140; GB:U00096; NID:q1786520; PIDN:AACT73436.1; PID:q1786527;
A:Experimental source: strain K-12, substrain M61655
R:Patton, A.J.; Hough, D.W.; Townner, P.; Danson, M.J.
Eur. J. Biochem. 214, 75-81, 1993
A:Title: Does Escherichia coli possess a second citrate synthase gene?
A:Reference number: S33286; MUID:93285168; PMID:8508809
A:Accession: S33286
A:Status: preliminary
A:Molecule type: protein
A:Residues: 'N', 19-32, 'X', 34-42, 'X', 44-47, 'XX', 50-52, 'X', 54 <PAT>
C:Genetics:
A:Gene: prbc
C:Function:
A:Description: catalyzes synthesis of citrate from acetyl-CoA and oxaloacetic acid
A:Pathway: tricarboxylic acid cycle
C:Superfamily: citrate (si)-synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 96.8%; Score 30; DB 2; Length 389;
Best Local Similarity 83.3%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 65 LVHGKL 70

RESULT 4
B90677
probable citrate synthase [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: B90677
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90677
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <HAY>
A:Cross-references: PIDN:BA033809.1; PID:q13359843; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0386
C:Superfamily: citrate (si)-synthase

Query Match 96.8%; Score 30; DB 2; Length 389;
Best Local Similarity 83.3%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 65 LVHGKL 70

RESULT 5
E85527
hypothetical protein prpC [imported] - Escherichia coli (strain O157:H7, substrain ED
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85527
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STO>
A:Cross-references: GB:AF005174; NID:q12513160; PIDN:AAG54681.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: prpC
C:Superfamily: citrate (si)-synthase

Query Match 96.8%; Score 30; DB 2; Length 389;
Best Local Similarity 83.3%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 65 LVHGKL 70

RESULT 6
AH0547
methylcitrate synthase (EC 4.1.3.-) [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0547
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08824.1; PID:q16501638; GSPDB:GN00176
C:Genetics:
A:Gene: prpC
C:Superfamily: citrate (si)-synthase
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 96.8%; Score 30; DB 2; Length 389;
Best Local Similarity 83.3%; Pred. No. 76;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
|:||||
Db 65 LVHGKL 70

RESULT 7
B42529
B-ORF-A protein - vaccinia virus (strains WR and Copenhagen)
C:Species: vaccinia virus
A:Note: host Homo sapiens (man)

C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 20-Sep-1999
C;Accession: B42529; JQ1827
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172

A;Accession: B42529
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-107 <JOH>
R;Smith, G.L.; Chan, Y.S.; Howard, S.T.
J. Gen. Virol. 72, 1349-1376, 1991

A;Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right in
A;Reference number: JQ1767; MUID:91259063; PMID:2045793

A;Accession: JQ1827
A;Molecule type: DNA
A;Residues: 1-107 <SMI>
A;Cross-references: DDBJ:D11079
C;Superfamily: vaccinia virus B-ORF-A protein

Query Match 93.5%; Score 29; DB 2; Length 107;
Best Local Similarity 83.3%; Pred. No. 32;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 43 LVHGKI 48

RESULT 8

T04091
hypothetical protein - maize

C;Species: Zea mays (maize)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C;Accession: T04091
R;Betawar, N.M.; Baysdorfer, C.

submitted to the EMBL Data Library, July 1996
A;Description: Novel maize gene.

A;Reference number: Z15202
A;Accession: T04091
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-143 <BET>

A;Cross-references: EMBL:U64437; NID:g1498054; PIDN:AAB06331.1; PID:g1498055
A;Experimental source: strain B73

Query Match 93.5%; Score 29; DB 2; Length 143;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 1 MVHGKL 6

RESULT 9

B84063

hypothetical protein BH3306 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B84063
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11059132
A;Accession: B84063
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <STO>

A;Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BAB07025.1; GSPDB:GN00

A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3306

Query Match 93.5%; Score 29; DB 2; Length 265;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 198 IVHGKL 203

RESULT 10

B82185

glycerol-3-phosphate ABC transporter, ATP-binding protein VC1552 [imported] - Vibrio

C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 19-Jul-2002
C;Accession: B82185

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: B82185

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-393 <HEI>

A;Cross-references: GB:AE004233; GB:AE003852; NID:g9656055; PIDN:AAF94706.1; GSPDB:G

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC1552

A;Map position: 1

C;Superfamily: inner membrane protein malk; ATP-binding cassette homology

Query Match 93.5%; Score 29; DB 2; Length 393;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 326 LVHGKI 331

RESULT 11

T30089

probable zinc proteinase (EC 3.4.24.-) C02G6.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T30089

R;Bentley, D.; Kemp, K.; Scheet, P.

submitted to the EMBL Data Library, April 1996

A;Description: The sequence of C. elegans cosmid C02G6.

A;Reference number: Z20734

A;Accession: T30089

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-980 <BEN>

A;Cross-references: EMBL:U55372; PIDN:AAA98001.1; GSPDB:GN00023; CESP:C02G6.1

A;Experimental source: strain Bristol N2; clone C02G6

C;Genetics:

A;Gene: CESP:C02G6.1

A;Map position: 5

A;Introns: 25/3; 215/2; 266/3; 540/3; 585/3; 786/1; 898/2

C;Superfamily: Insulysin

C;Keywords: hydrolase; metalloproteinase; zinc

F;70/74/Binding site: zinc (His) #status predicted

F;73/Active site: Glu #status predicted

Query Match 93.5%; Score 29; DB 2; Length 980;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 913 LVHGKI 918

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: D97260
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <KUR>
 A:Cross-references: GB:AF001437; PIDN:AAK80871.1; PID:gl5025980; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2929

Query Match 90.3%; Score 28; DB 2; Length 188;
 Best Local Similarity 83.3%; Pred. No. 97;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 136 VVHGKL 141

RESULT 15
 F84061
 hypothetical protein BH3294 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Dec-2002
 C:Accession: F84061
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: F84061
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-254 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:gl0175792; PIDN:BA07013.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3294
 C:Superfamily: uncharacterized conserved protein

Query Match 90.3%; Score 28; DB 2; Length 254;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 245 VVHGKL 250

RESULT 16
 T01758
 hypothetical protein A_IC002P16.1 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 22-Oct-1999
 C:Accession: T01758
 R;Miller, N.; Beck, C.; Kramer, J.
 submitted to the EMBL Data Library, June 1997
 A:Description: The sequence of A. thaliana IC002P16.
 A:Reference number: Z14421
 A:Accession: T01758
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-287 <MIL>
 A:Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191158; GSPDB:GN00063; ATSP:A
 C:Genetics:
 A:Gene: ATSP:A_IC002P16.1
 A:Map position: 5
 A:Introns: 167/3

Query Match 90.3%; Score 28; DB 2; Length 287;

RNA polymerase beta chain [imported] - Nostoc sp. (strain PCC 7120)
 A:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD2005
 R;Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriugu
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2005
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1131 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAH77960.1; PID:gl7135414; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: rpoB
 C:Superfamily: DNA-directed RNA polymerase beta chain

Query Match 93.5%; Score 29; DB 2; Length 1131;
 Best Local Similarity 83.3%; Pred. No. 4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 931 IVHGKL 936

RESULT 13
 A86370
 hypothetical protein F508.17 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
 C:Accession: A86370
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.B.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: A86370
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-178 <STO>
 A:Cross-references: GB:AE005172; NID:g4056444; PIDN:AAC98017.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 90.3%; Score 28; DB 2; Length 178;
 Best Local Similarity 83.3%; Pred. No. 91;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 7 LLHGKL 12

RESULT 14
 D97260
 probable membrane protein [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: D97260

Query Match 90.3%; Score 28; DB 2; Length 287;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 67 LVHGRL 72

RESULT 17

JC5644
acyl-CoA thiolesterase (EC 3.1.2.-) III, peroxisomal - human
N:Alternate names: HIV Nef-binding protein
N:Contains: palmitoyl-CoA hydrolase (EC 3.1.2.2)
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-May-2000
C:Accession: JC5644
R:Watanabe, H.; Shiratori, T.; Shoji, H.; Miyatake, S.; Okazaki, Y.; Ikuta, K.; Sato, T.
Biochem. Biophys. Res. Commun. 238, 234-239, 1997
A:Title: A novel acyl-CoA thioesterase enhances its enzymatic activity by direct binding
A:Reference number: JC5644; MUID:97445158; PMID:9299485
A:Accession: JC5644
A:Molecule type: mRNA
A:Residues: 1-319 <WAT>
A:Cross-references: DDBJ:AF014404; NID:g2318124; PIDN:AAB71665.1; PID:g2318125
C:Comment: This enzyme regulates membrane sorting and subcellular morphology.
C:Genetics:
A:Gene: GDB:PTE1
A:Cross-references: GDB:9954875
C:Superfamily: acyl-CoA thiolesterase II
C:Keywords: coenzyme A; peroxisome; thiolester hydrolase
F:317-319/Region: peroxisome/glyoxysome location signal (S-[RKH]-L) motif
F:78/Active site: His #status predicted

Query Match 90.3%; Score 28; DB 1; Length 319;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 286 LVHGRL 291

RESULT 18

H64396
hypothetical protein M0776 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: H64396
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: H64396
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-383 <BUL>
A:Cross-references: GB:U67522; GB:L77117; NID:g2826315; PIDN:AAB98766.1; PID:g1499596; T
C:Genetics:
A:Map position: FOR697377-698528

Query Match 90.3%; Score 28; DB 2; Length 383;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 65 LVHGRL 70

RESULT 19

C:Species: Chlamydia pneumoniae [imported] - Chlamydia pneumoniae (strain J138)

S23817

hypothetical protein Tnpl - garden snapdragon transposable element Tnpl
C:Species: Antirrhinum majus (garden snapdragon)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000
C:Accession: S78587; S23817
R:Nacken, W.K.F.; Piotrowiak, R.; Saedler, H.; Sommer, H.
Mol. Gen. Genet. 228, 201-208, 1991
A:Title: The transposable element Tnpl from Antirrhinum majus shows structural homology
A:Reference number: S16551; MUID:91360065; PMID:1715971
A:Accession: S78587
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-484 <NAC>
A:Cross-references: EMBL:X57297; NID:g16056; PIDN:CAA40554.1; PID:g16057
C:Genetics:

A:Molecule type: transposable element Tnpl
A:Introns: 11/3; 57/1; 139/3; 206/3; 237/3; 271/3; 330/3; 468/2
C:Superfamily: garden snapdragon transposable element Tnpl hypothetical protein Tnpl

Query Match 90.3%; Score 28; DB 2; Length 484;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 355 LVHGKM 360

RESULT 20

C72002
exodeoxyribonuclease, large chain CP0787 [imported] - Chlamydia pneumoniae (strain
N:Alternate names: exodeoxyribonuclease vii
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 02-Sep-2000
C:Accession: C72002; H81538
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-554 <ARN>
A:Cross-references: GB:AE001686; GB:AE001363; NID:g4377389; PIDN:RAD19199.1; PID:g437
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
; C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: H81538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-554 <REA>
A:Cross-references: GB:AE002238; GB:AE002161; NID:g7189693; PIDN:AAF38586.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: xseA; CP0787
C:Superfamily: exodeoxyribonuclease VII

Query Match 90.3%; Score 28; DB 2; Length 554;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 77 LVHGKL 82

RESULT 21

C86623

C:Species: Chlamydia pneumoniae [imported] - Chlamydia pneumoniae (strain J138)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C;Accession: C86623
 R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
 Nucleic Acids Res. 28, 2311-2314, 2000
 A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A;Reference number: A86491; MUID:20330349; PMID:10871362
 A;Accession: C86623
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-554 <STO>
 A;Cross-references: GB:BA000008; NID:g8979435; PIDN:BAA99269.1; GSPDB:GN00142
 A;Experimental source: strain J138
 C;Genetics:
 A;Gene: xsaA
 C;Superfamily: exodeoxyribonuclease VII

Query Match 90.3%; Score 28; DB 2; Length 554;
 Best Local Similarity 66.7%; Pred. No. 3.1e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LVHGKL 6
 :|||||
 Db 77 IIHGKL 82

RESULT 22
 T19839
 hypothetical protein C39B10.1 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C;Accession: T19839
 R;White, S.
 submitted to the EMBL Data Library, July 1995
 A;Reference number: Z19186
 A;Accession: T19839
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-583 <WTL>
 A;Cross-references: EMBL:Z50027; PIDN:CAA90331.1; GSPDB:GN00028; CESP:C39B10.1
 A;Experimental source: clone C39B10
 C;Genetics:
 A;Gene: CESP:C39B10.1
 A;Map position: X
 A;Introns: 43/2; 68/3; 99/1; 190/3; 226/1; 278/3; 307/2; 346/3; 390/3; 529/3

Query Match 90.3%; Score 28; DB 2; Length 583;
 Best Local Similarity 66.7%; Pred. No. 3.2e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 LVHGKL 6
 :|||||
 Db 401 IIHGKL 406

RESULT 23
 H69879
 ATP-dependent DNA helicase homolog ylpB - Bacillus subtilis
 C;Species: Bacillus subtilis
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C;Accession: H69879
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth
 C.; Bron, S.; Brouillet, S.; Bruschii, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: H69879
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-682 <GUN>
 A;Cross-references: GB:Z99112; GB:AL009126; NID:g26333902; PIDN:CAB13460.1; PID:g26333
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: ylpB
 C;Superfamily: DNA helicase recG
 C;Keywords: ATP; nucleotide binding; P-loop
 F;284-291/Region: nucleotide-binding motif A (P-loop)
 F;381-386/Region: nucleotide-binding motif B
 F;385-388/Region: DEXH motif

Query Match 90.3%; Score 28; DB 1; Length 682;
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 506 LMHGKL 511

RESULT 24

AC1301
 ATP-dependent DNA helicase recG homolog lmo1811 [imported] - Listeria monocytogenes (C;
 C;Species: Listeria monocytogenes
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C;Accession: AC1301
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A;Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AC1301
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-682 <GIA>
 A;Cross-references: GB:NC_003210; PIDN:CAC99889.1; PID:gi6411265; GSPDB:GN00177
 A;Experimental source: strain EGD-e
 C;Genetics:
 A;Gene: lmo1811
 C;Superfamily: DNA helicase recG

Query Match 90.3%; Score 28; DB 2; Length 682;
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 506 LMHGKL 511

RESULT 25

AC1673
 ATP-dependent DNA helicase recG homolog lin1925 [imported] - Listeria innocua (strain
 C;Species: Listeria innocua
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C;Accession: AC1673
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A;Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AC1673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-682 <GLA>
A:CROSS-references: GB:AL592022; PIDN:CAC97155.1; PID:g161414426; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin1925
C:Superfamily: DNA helicase recG

Query Match 90.3%; Score 28; DB 2; Length 682;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||
Db 506 LMHGKL 511

RESULT 26
F89895
ATP-dependent DNA helicase [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89895
R:Ruroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89895
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-686 <KUR>
A:CROSS-references: GB:BA000018; PID:g13701027; PIDN:BAB42322.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: recG
C:Superfamily: DNA helicase recG

Query Match 90.3%; Score 28; DB 2; Length 686;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||
Db 513 LMHGKL 518

RESULT 27
I65413
sodium-dependent neurotransmitter transporter - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 20-Aug-1999
C:Accession: I65413
R:Uhl, G.R.; Kitayama, S.; Gregor, P.; Nanthakumar, E.; Persico, A.; Shimada, S.
Brain Res. Mol. Brain Res. 16, 353-359, 1992
A:Title: Neurotransmitter transporter family cDNAs in a rat midbrain library: 'orphan'
A:Reference number: I52632; MUID:93180651; PMID:1363329
A:Accession: I65413
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-728 <RES>
A:CROSS-references: GB:S56968; NID:g298323; PIDN:AAB25532.1; PID:g298324
C:Superfamily: gamma-aminobutyric acid transporter

Query Match 90.3%; Score 28; DB 2; Length 728;
Best Local Similarity 66.7%; Pred. No. 4.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||

Db 675 LIHGKI 680

RESULT 28
I52632
sodium-dependent neurotransmitter transporter - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 21-Aug-1998
C:Accession: I52632
R:Uhl, G.R.; Kitayama, S.; Gregor, P.; Nanthakumar, E.; Persico, A.; Shimada, S.
Brain Res. Mol. Brain Res. 16, 353-359, 1992
A:Title: Neurotransmitter transporter family cDNAs in a rat midbrain library: 'orphan'
A:Reference number: I52632; MUID:93180651; PMID:1363329
A:Accession: I52632
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-730 <RES>
A:CROSS-references: GB:S56141; NID:g298325
C:Superfamily: gamma-aminobutyric acid transporter

Query Match 90.3%; Score 28; DB 2; Length 730;
Best Local Similarity 66.7%; Pred. No. 4.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|:|||||
Db 677 LIHGKI 682

RESULT 29
S48521
AKR1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: ankyrin repeat-containing protein; protein YD9230B.03c; protein YD
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 20-Sep-1999
C:Accession: S48521; S61119; S67467; S70124; JC6085
R:Bender, A.; Peterson, J.
submitted to the EMBL Data Library, April 1994
A:Description: Characterization of the yeast ankyrin repeat-containing protein Akrlp.
A:Reference number: S48521
A:Accession: S48521
A:Molecule type: DNA
A:Residues: 1-764 <BEN>
A:CROSS-references: EMBL:L31407; NID:g466521; PIDN:AAC41676.1; PID:g466522
R:Murphy, L.; Harris, D.
submitted to the EMBL Data Library, December 1995
A:Reference number: S61117
A:Accession: S61119
A:Molecule type: DNA
A:Residues: 674-764 <MUR>
A:CROSS-references: EMBL:268290; NID:g1134887; PIDN:CAA92582.1; PID:e214677; PID:g113
R:Murphy, L.; Harris, D.
submitted to the EMBL Data Library, March 1996
A:Reference number: S67454
A:Accession: S67467
A:Molecule type: DNA
A:Residues: 674-764 <MUR>
A:CROSS-references: EMBL:270202; NID:g1226026; PIDN:CAA94103.1; PID:e228632; PID:g122
R:Le, T.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of S. cerevisiae cosmid 9954.
A:Reference number: S70124
A:Accession: S70124
A:Molecule type: DNA
A:Residues: 1-764 <LET>
A:CROSS-references: EMBL:U51030; NID:g1332633; PIDN:AAB64454.1; PID:g1230637; MIPS:YD
R:Kao, L.R.; Peterson, J.; Ji, R.; Bender, L.; Bender, A.
Mol. Cell. Biol. 16, 168-178, 1996
A:Title: Interactions between the ankyrin repeat-containing protein Akrlp and the phe
A:Reference number: JC6085; MUID:96104566; PMID:8524293
A:Accession: JC6085
A:Molecule type: DNA
A:Residues: 1-764 <KAO>

A;Cross-references: GB:L31407; NID:g466521; PIDN:AAC41676.1; PID:g466522
C;Genetics:
A;Gene: SGD:AKR1
A;Cross-references: MIPS:YDR264c; SGD:S0002672
A;Map position: 4R
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
C;Keywords: transmembrane protein
F:189-205/Domain: transmembrane #status predicted <TM1>
F:324-340/Domain: transmembrane #status predicted <TM2>
F:344-360/Domain: transmembrane #status predicted <TM3>
F:387-403/Domain: transmembrane #status predicted <TM4>
F:422-438/Domain: transmembrane #status predicted <TM5>
F:516-532/Domain: transmembrane #status predicted <TM6>
F:571-587/Domain: transmembrane #status predicted <TM7>

Query Match 90.3%; Score 28; DB 2; Length 764;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 92 MIHGKL 97

RESULT 30
T49744
probable ubiquitin-protein ligase [imported] - Neurospora crassa
N:Alternate names: protein B24B19.160
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C;Accession: T49744
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49744
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-815 <SCH>
A;Cross-references: EMBL:AL356192; GSPDB:GN001116; NCSP:B24B19.160
A;Experimental source: BAC clone B24B19; strain OR74A
C;Genetics:
A;Gene: NCSP:B24B19.160
A;Map position: 6
A;Intons: 11/1; 24/1; 59/2; 110/1; 783/2
C;Superfamily: yeast ubiquitin-protein ligase; WW repeat homology
F:239-276/Domain: WW repeat homology <WWR1>
F:334-371/Domain: WW repeat homology <WWR2>
F:393-430/Domain: WW repeat homology <WWR3>

Query Match 90.3%; Score 28; DB 2; Length 815;
Best Local Similarity 83.3%; Pred. No. 4.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 125 VVHGKL 130

RESULT 31
AH0437
Clp ATPase [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AH0437
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0437
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-867 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92828.1; PID:g15981519; GSPDB:GN00175
C;Genetics:
A;Gene: clpB2
C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 90.3%; Score 28; DB 2; Length 867;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 777 IIHGKL 782

RESULT 32
A89201
protein F32D8.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: A89201
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: A89201
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-870 <STO>
A;Cross-references: GB:chr_V; PIDN:CAA98456.1; PID:g3876615; GSPDB:GN00023; CESP:F32D
C;Genetics:
A;Gene: F32D8.4
A;Map position: 5

Query Match 90.3%; Score 28; DB 2; Length 870;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 144 LLHGKL 149

RESULT 33
A46764
microsomal triglyceride transfer protein - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 17-May-1996 #sequence_revision 07-Nov-1997 #text_change 05-Nov-1999
C;Accession: A46764
R;Sharp, D.; Blinderman, L.; Combs, K.A.; Klenzle, B.; Ricci, B.; Wager-Smith, K.; Gi
Nature 365, 65-69, 1993
A;Title: Cloning and gene defects in microsomal triglyceride transfer protein associa
A;Reference number: S36757; MUID:93368660; PMID:8361539
A;Accession: A46764
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-887 <SHA>
A;Cross-references: EMBL:X78567; NID:g469090; PIDN:CAA55310.1; PID:g469091
C;Comment: This protein forms a heterodimer with protein disulfide isomerase.

Query Match 90.3%; Score 28; DB 2; Length 887;
Best Local Similarity 66.7%; Pred. No. 5.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
:|||||
Db 120 LIHGKI 125

RESULT 34
T21659

hypothetical protein F32D8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21659
R:Wilkinson, J.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19454

A:Accession: T21659

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-912 <WIL>

A:Cross-references: EMBL:Z74031; PIDN:CAA98456.2; GSPDB:GN00023; CESP:F32D8.4

A:Experimental source: clone F32D8

C:Genetics:

A:Map position: 5

A:Introns: 15/1; 42/1; 75/1; 243/3; 304/3; 406/3; 444/2; 486/2; 581/1; 662/2; 787

Query Match 90.3%; Score 28; DB 2; Length 912;

Best Local Similarity 83.3%; Pred. No. 5.2e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 144 LLHGKL 149

RESULT 35

F70177

transcription-repair coupling factor (mfd) homolog - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: F70177

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: F70177

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1125 <KLE>

A:Cross-references: GB:AE001163; GB:AE000783; NID:g2688541; PIDN:AAC66973.1; PID:g268854

A:Experimental source: strain B31

C:Superfamily: transcription-repair coupling protein

C:Keywords: ATP; DNA repair; leucine zipper; nucleotide binding; P-loop; transcription

F;610-617/Region: nucleotide-binding motif A (P-loop)

F;707-712/Region: nucleotide-binding motif B

F;711-714/Region: DEAD/H motif #status atypical

Query Match

Best Local Similarity 90.3%; Score 28; DB 1; Length 1125;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 822 IIHGKL 827

RESULT 36

T31657

reverse transcriptase homolog - sea squirt (Clona intestinalis)

C:Species: Clona intestinalis

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T31657

R:Bird, A.P.; Clark, V.; Jones, S.J.; Leighteb, S.; Dobson, R.; Tweedie, S.

submitted to the EMBL Data Library, December 1996

A:Reference number: Z21049

A:Accession: T31657

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1209 <BIR>

A:Cross-references: EMBL:Z83760; NID:e1014349; PID:e289993; PIDN:CAB06047.1

C:Genetics:

A:Note: COS41.3

Query Match

Best Local Similarity 90.3%; Score 28; DB 2; Length 1209;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 1035 LLHGKL 1040

RESULT 37

S67483

adenosinetriphosphatase 2 - malaria parasite (Plasmodium falciparum)

N:Alternate names: ATPase 2

C:Species: Plasmodium falciparum

C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jun-2000

C:Accession: S67483

R:Prottein, F.; Cowman, A.F.

Eur. J. Biochem. 227, 214-225, 1995

A:Title: Molecular cloning and sequence of two novel P-type adenosinetriphosphatases

A:Reference number: S67483; MUID:95154293; PMID:7851389

A:Accession: S67483

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1553 <TRO>

A:Cross-references: EMBL:U16955

C:Genetics:

A:Introns: 17/2

Query Match

Best Local Similarity 90.3%; Score 28; DB 2; Length 1553;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 1258 LVHGKL 1263

RESULT 38

T21588

hypothetical protein F31C3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21588

R:Cottage, A.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19446

A:Accession: T21588

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2025 <WIL>

A:Cross-references: EMBL:Z92784; PIDN:CAB07193.1; GSPDB:GN00019; CESP:F31C3.3

A:Experimental source: clone F31C3

C:Genetics:

A:Gene: CESP:F31C3.3

A:Map position: 1

A:Introns: 395/2; 549/2; 768/1; 925/1; 1076/3; 1871/3; 1927/3; 2006/3

Query Match

Best Local Similarity 90.3%; Score 28; DB 2; Length 2025;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6

Db 971 LVHGKL 976

RESULT 39

D42088

adenylate cyclase (EC 4.6.1.1), Ca2+/calmodulin-responsive - fruit fly (*Drosophila melanogaster*)
N:Alternate names: rutabaga protein
C:Species: *Drosophila melanogaster*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: D42088
R:Levin, L.R.; Han, P.L.; Hwang, P.M.; Feinstein, P.G.; Davis, R.L.; Reed, R.R.
Cell 68, 479-489, 1992
A:Title: The *Drosophila* learning and memory gene *rutabaga* encodes a Ca2+/calmodulin-responsive protein
A:Reference number: A42088; MUID:92154664; PMID:1739965
A:Accession: D42088
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2248 <LEV>
A:Cross-references: GB:M81887; NID:g158194; PIDN:AAA28844.1; PID:g158195
A:Note: sequence extracted from NCBI backbone (NCBIP:82229)
C:Genetics:
A:Gene: FlyBase: rut
A:Cross-references: FlyBase:FBgn0003301
C:Superfamily: *Drosophila* adenylate cyclase; guanylate cyclase catalytic domain homology
C:Keywords: phosphorus-oxygen lyase; transmembrane protein
F:218-452/Domain: guanylate cyclase catalytic domain homology <GCC1>
F:910-1155/Domain: guanylate cyclase catalytic domain homology <GCC2>

Query Match 90.38; Score 28; DB 1; Length 2248;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
DB 893 LVHGRL 898

RESULT 40
C84085
hypothetical protein BH3483 [imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84085
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and its relationship to other members of the family Halobacteriaceae
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84085
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-78 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07202.1; GSPDB:GN001186
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3483

Query Match 87.1%; Score 27; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
|||||
DB 12 VHGKL 16

RESULT 41
A97550
hypothetical protein AGR_C_2893 [imported] - *Agrobacterium tumefaciens* (strain C58, Cerevisiae)
C:Species: *Agrobacterium tumefaciens*
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: A97550
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* strain C58
A:Reference number: A97550; MUID:21608551; PMID:11743194
A:Accession: A97550
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-105 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87354.1; PID:g15156658; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2893
A:Map position: circular chromosome

Query Match 87.1%; Score 27; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|||||
DB 5 LVHGK 9

RESULT 42

AH2769
hypothetical protein Atul570 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)
C:Species: *Agrobacterium tumefaciens*
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2769
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wooledge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClintock, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamster, E.W.
A:Title: The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2769
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAI42574.1; PID:g17739999; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul570
A:Map position: circular chromosome

Query Match 87.1%; Score 27; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|||||
DB 5 LVHGK 9

RESULT 43

H70867
hypothetical protein Rv2481c - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70867
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon-Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70867
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-107 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16058.1; PID:e123
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2481c

Query Match 87.1%; Score 27; DB 2; Length 107;
Best Local Similarity 83.3%; Pred. No. 88;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 79 LVHQL 84

RESULT 44

C87439
conserved hypothetical protein CCI532 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87439
R:NIERMAN, W.C.; FEIDBYLUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON
N, J.; ERMOLAeva, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87439
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-124 <STO>
A:Cross-references: GB:AE005673; NID:g13422915; PIDN:AAK23511.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCI532

Query Match 87.1%; Score 27; DB 2; Length 124;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
|||||
Db 17 VHGKL 21

RESULT 45

C69061
hypothetical protein MTH1459 - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69061
R:SMITH, D.R.; DOUCETTE-STAMM, L.A.; DELOUGHRY, C.; LEE, H.; DUBOIS, J.; ALDREDGE, T.;
QIU, D.; SPADAFORA, R.; VICAIRe, R.; WANG, Y.; WIERZBOWSKI, J.; GIBSON, R.; JIWANI, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: C69061
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-130 <MTH>
A:Cross-references: GB:AE000907; GB:AE000666; NID:g2622568; PIDN:AA85934.1; PID:g262257
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1459
A:Start codon: GTG

Query Match 87.1%; Score 27; DB 2; Length 130;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
|||||
Db 114 LVHGK 118

Search completed: October 14, 2003, 13:12:54

Job time : 16.1818 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 7.63636 Seconds
(without alignments)
36.950 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29
Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	96.8	216	ME13_SCHPO	Q9hgk2 schizosacch
2	30	96.8	389	PRPC_ECOLI	P31860 escherichia
3	30	96.8	389	PRPC_SALTY	Q56063 salmonella
4	29	93.5	107	YVBA_VACCV	P20541 vaccinia vi
5	29	93.5	1131	RPOB_ANASP	P22703 anabaena sp
6	28	90.3	319	PTEL_HUMAN	Q14734 homo sapien
7	28	90.3	320	PTEL_MOUSE	P58137 mus musculu
8	28	90.3	383	Y776_METJA	Q58186 methanococc
9	28	90.3	554	EX7L_CHLPN	Q92677 chlamydia p
10	28	90.3	682	RECG_BACSU	Q34942 bacillus su
11	28	90.3	686	RECG_STAAD	O50581 staphylococ
12	28	90.3	729	NTW7_BOVIN	Q9xs59 bos taurus
13	28	90.3	729	NTW7_RAT	Q08469 rattus norv
14	28	90.3	730	NTW7_HUMAN	Q9hzj7 homo sapien
15	28	90.3	764	AKRL_YEAST	P39010 saccharomyc
16	28	90.3	887	MTF_BOVIN	P55156 bos taurus
17	28	90.3	1125	MFD_BORBU	O51568 borrelia bu
18	28	90.3	2248	CYAL_DROME	P32870 drosophila
19	27	87.1	184	RHEB_HUMAN	Q15382 homo sapien
20	27	87.1	184	RHEB_MOUSE	Q121j2 mus musculu
21	27	87.1	184	RHEB_RAT	Q62839 rattus norv
22	27	87.1	195	HIS7_THEMA	Q9tdx9 thermotoga
23	27	87.1	215	TCPL_HUMAN	Q8tdr4 homo sapien
24	27	87.1	223	SRV_HORSE	P36389 equus cabal
25	27	87.1	239	SFSA_CAUCR	Q9a4z6 caulobacter
26	27	87.1	246	NAGB_BACHD	Q9kf98 bacillus ha
27	27	87.1	304	COVD_BACSU	P39776 bacillus su
28	27	87.1	310	EXDI_ORISA	Q9awm9 oryza sativ
29	27	87.1	312	RLUD_BUCAI	P57481 buchnera ap
30	27	87.1	316	MCH_ARCFU	Q28344 archaeoglob
31	27	87.1	334	G3PL_BACSU	P09124 bacillus su
32	27	87.1	334	G3PL_BACME	P23722 bacillus me
33	27	87.1	344	ADK1_ARATH	Q9sf85 arabidopsis

RESULT 1
ME13_SCHPO

ID	ME13_SCHPO	STANDARD;	PRT;	216 AA.
AC	O9HGK2;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Homologous pairing protein meul3.			
GN	MEU13 OR SPAC222.15 OR SPAC821.01.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.			
RX	MEDLINE=21340279; PubMed=11447128;			
RA	Nabeshima K., Kakihara Y., Hiraoka Y., Nojima H.;			
RT	'A novel meiosis-specific protein of fission yeast, Meul3p, promotes			
RL	homologous pairing independently of homologous recombination.';			
RL	EMBO J. 20:3871-3881(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CD16-1;			
RX	MEDLINE=21270454; PubMed=11376151;			
RA	Watanabe T., Miyashita K., Saito T.T., Yoneki T., Kakihara Y.,			
RA	Nabeshima K., Kishi Y.A., Shimoda C., Nojima H.;			
RT	'Comprehensive isolation of meiosis-specific genes identifies novel			
RT	proteins and unusual non-coding transcripts in Schizosaccharomycetes			
RT	'Pombe.';			
RL	Nucleic Acids Res. 29:2327-2337(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE=21848401; PubMed=11859360;			
RA	Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volckaert G., Aert R., Robben J., Grynponez B.,			
RA	Weltjens I., Vanscreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			

P55318 homo sapien
P40990 saccharomyc
P35584 mus musculu
P32183 rattus norv
P71348 haemophilus
Q8yjf9 b arginine
Q8fve2 b arginine
P25809 rattus norv
Q28316 archaeoglob
P57221 buchnera ap
P44104 haemophilus
Q9zmp8 helicobacte

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of *Schizosaccharomyces pombe*.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Required for proper homologous pairing and efficient
CC cross-over and intragenic recombination during meiosis. Acts
CC indirectly in a process facilitating homologous recombination. Acts
CC during mid- to late-horse-tail period.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017038; BAB17055.1; -;
CC EMBL; AL132798; CAD33796.1; -;
CC EMBL; AL121770; CAD33805.1; -;
CC GeneDB.SPombe; SPAC222.15; -;
CC MELOSIS; DNA recombination; Nuclear protein.
CC SEQUENCE 216 AA; 24672 MW; ESA3839FAC13340A CRC64;
CC -----
CC Query Match 96.8%; Score 30; DB 1; Length 216;
CC Best Local Similarity 83.3%; Pred. No. 11;
CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1 LVHGKL 6
CC I:|||||
CC DB 62 LIHGKL 67

RESULT 2
PRPC_ECOLI STANDARD; PRT; 389 AA.
ID PRPC_ECOLI STRAIN-K12 / MG1655;
AC P31660; P77217;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 2-methylcitrate synthase (EC 2.3.3.5) (Methylcitrate synthase)
DE (Citrate synthase 2).
DE PRPC OR B0333.
OS *Escherichia coli*.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE OF 18-54, AND CHARACTERIZATION.
RX MEDLINE=93285168; PubMed=8508809;
RA Patton A.J., Hough D.W., Towner P., Danson M.J.;
RT "Does *Escherichia coli* possess a second citrate synthase gene?";
RL Eur. J. Biochem. 214:75-81(1993).
RN [4]
RP CHARACTERIZATION.

RX MEDLINE=98004569; PubMed=9325432;
RA Textor S., Wendisch V.F., de Graaf A.A., Mueller U., Linder M.I.,
RA Linder D., Buckel W.;
RT "Propionate oxidation in *Escherichia coli*: evidence for operation of a
RT methylcitrate cycle in bacteria.";
RL Arch. Microbiol. 168:428-436(1997).
CC -!- FUNCTION: Catalyzes the synthesis of 2-methylcitrate from
CC propionyl-CoA and oxaloacetate. Also catalyzes the condensation of
CC oxaloacetate with acetyl-CoA but with a lower specificity.
CC -!- CATALYTIC ACTIVITY: Propionyl-CoA + H(2)O + oxaloacetate =
CC (2R,3S)-2-hydroxybutane-1,2,3-tricarboxylate + COA.
CC -!- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; second
CC step.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SIMILARITY: Belongs to the citrate synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000140; AAC73436.1; -;
CC EMBL; U73857; AAB18057.1; -;
CC PIR; E64760; E64760.
CC HSSP; O34002; 1A59.
CC Ecogene; Egi1756; prpc.
CC InterPro; IPR002020; Citrate_synt.
CC Pfam; PF00285; citrate_synt; 1.
CC PRINTS; PR00143; CITRATNTSHASE.
CC PROSITE; PS00480; CITRATE_SYNTHASE; 1.
CC Transfaser; Complete proteome.
CC ACT_SITE 274 274 BY SIMILARITY.
CC ACT_SITE 325 325 BY SIMILARITY.
CC ACT_SITE 18 18 K -> N (IN REF. 2).
CC CONFLICT 18 18
CC SEQUENCE 389 AA; 43102 MW; 02B779E7AD4581C3 CRC64;
CC -----
CC Query Match 96.8%; Score 30; DB 1; Length 389;
CC Best Local Similarity 83.3%; Pred. No. 21;
CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1 LVHGKL 6
CC I:|||||
CC DB 65 LIHGKL 70

RESULT 3
PRPC_SALTY STANDARD; PRT; 389 AA.
ID PRPC_SALTY STANDARD; PRT; 389 AA.
AC Q56063;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 2-methylcitrate synthase (EC 2.3.3.5) (Methylcitrate synthase)
DE (Citrate synthase 2).
DE PRPC OR STM0369.
GN *Salmonella typhimurium*.
OS *Salmonella typhimurium*.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; *Salmonella*.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=97158691; PubMed=9006051;
RA Horswill A.R., Escalante-Semerena J.C.;
RT "Propionate catabolism in *Salmonella typhimurium* LT2: two divergently
RT transcribed units comprise the prp locus at 8.5 centisomes, prpR
RT encodes a member of the sigma-54 family of activators, and the
RT prpBCDE genes constitute an operon.";
RL J. Bacteriol. 179:928-940(1997).
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SCSCI1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RT Nature 413:852-856(2001).
RN [3]
RP FUNCTION, AND PATHWAY.
RX MEDLINE=99412259; PubMed=10482501;
RA Horswill A.R., Escalante-Semerena J.C.;
RA "Salmonella typhimurium LT2 catabolizes propionate via the
RT 2-methylcitric acid cycle.";
RL J. Bacteriol. 181:5615-5623(1999).
CC -1- FUNCTION: Catalyzes the synthesis of 2-methylcitrate from
CC propionyl-CoA and oxaloacetate. Also catalyzes the condensation of
CC oxaloacetate with acetyl-CoA but with a lower specificity.
CC -1- CATALYTIC ACTIVITY: Propanoyl-CoA + H(2)O + oxaloacetate =
CC (2R,3S)-2-hydroxybutane-1,2,3-tricarboxylate + CoA.
CC -1- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; second
CC step.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- SIMILARITY: Belongs to the citrate synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U51879; AAC44815.1; -.
CC DR EMBL; AE008712; AL019323.1; -.
CC DR HSP; O34002; 1A59.
CC DR StyGene; SGI0687; prpc.
CC DR InterPro; IPR002020; Citrate_synt.
CC DR Pfam; PF00285; citrate_synt; 1.
CC DR PRINTS; PR00143; CITRATESYNTHASE.
CC DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
CC Transferase; Complete proteome.
CC ACT_SITE 274 274 BY SIMILARITY.
CC ACT_SITE 325 325 BY SIMILARITY.
CC CONFLICT 144 144 S -> N (IN REF. 1).
CC SEQUENCE 389 AA; 43173 MW; 0927008F1E5F38D4 CRC64;
CC
CC Query Match 96.8%; Score 30; DB 1; Length 389;
CC Best Local Similarity 83.3%; Pred. No. 21;
CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 LVHGKL 6
CC Db 65 LIHGKL 70
CC
CC RESULT 4
CC YVBA_VACCV STANDARD; PRT; 107 AA.
CC AC P20541;
CC DT 01-FEB-1991 (Rel. 17, Created)
CC DT 01-FEB-1991 (Rel. 17, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hypothetical 11.8 kDa protein.
CC GN B ORF A.
CC OS Vaccinia virus (strain WR), and
CC OS Vaccinia virus (strain Copenhagen).
CC OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
CC Orthopoxvirus.
CC NCBI_TaxID=10254, 10249;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR;
RX MEDLINE=91259063; PubMed=2045793;
RA Smith G.L., Chan Y.S., Howard S.F.;
RA "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat.";
RL J. Gen. Virol. 72:1349-1376(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Copenhagen;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RA "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [3]
RP COMPLETE GENOME.
RC STRAIN=Copenhagen;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RA "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35027; AAA48195.1; -.
CC DR PIR; B42529; B42529.
CC KW Hypothetical protein.
CC SEQUENCE 107 AA; 11813 MW; FF9B5C8C1965F55F CRC64;
CC
CC Query Match 93.5%; Score 29; DB 1; Length 107;
CC Best Local Similarity 83.3%; Pred. No. 9.3;
CC Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 LVHGKL 6
CC Db 43 LVHGKI 48
CC
CC RESULT 5
CC RPOB_ANASP STANDARD; PRT; 1131 AA.
CC AC P22703;
CC DT 01-AUG-1991 (Rel. 19, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
CC beta chain) (RNA polymerase beta subunit).
CC GN RPOB OR ALR1594.
CC OS Anabaena sp. (strain PCC 7120).
CC OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CC NCBI_TaxID=103690;
CC
CC SEQUENCE FROM N.A.
CC RX MEDLINE=21595285; PubMed=11759840;
CC RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
CC RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
CC RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
CC RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
CC RA Yasuda M., Tabata S.;
CC "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 967-1131 FROM N.A.
RX MEDLINE=91258327; PubMed=1904436;

```

RA Bergsland K.J., Haselkorn R.;
 RT "Evolutionary relationships among eubacteria, cyanobacteria, and
 RT chloroplasts: evidence from the rpoC1 gene of Anabaena sp. strain PCC
 RT 7120.";
 RL J. Bacteriol. 173:3446-3455(1991).
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: IN CYANOBACTERIA THE RNA POLYMERASE IS COMPOSED OF FOUR
 CC SUBUNITS: ALPHA, BETA, GAMMA, AND DELTA.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AP003586; BAB77960.1; -;
 CC EMBL: M60831; AAA22032.1; -;
 CC PIR: A42361; A42361.
 CC PIR: AD2005; AD2005.
 CC HSP: Q9KWU7; 1HQW.
 CC InterPro: IPR001572; RNA_pol.B.
 CC Pfam: PF04563; RNA_pol_Rpb2_1; 1.
 CC Pfam: PF04561; RNA_pol_Rpb2_2; 1.
 CC Pfam: PF04565; RNA_pol_Rpb2_3; 1.
 CC Pfam: PF00562; RNA_pol_Rpb2_6; 1.
 CC Pfam: PF04560; RNA_pol_Rpb2_7; 1.
 CC PROSITE: PS01166; RNA_POL_BETA: 1.
 KW Transfrase: Transcription; DNA-directed RNA polymerase;
 KW Complete proteome.
 SQ SEQUENCE 1131 AA; 126524 MW; 7AB6EID28199FFD9 CRC64;
 Query Match 93.5%; Score 29; DB 1; Length 1131;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 :|||||
 Db 931 IVHGKL 936
 RESULT 6
 PTEI_HUMAN
 ID PTEI_HUMAN STANDARD; PRT; 319 AA.
 AC 014734; 015261;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peroxisomal long-chain acyl-coenzyme A thioester hydrolase 1 (EC 3.1.2.2)
 DE (Peroxisomal long-chain acyl-coA thioesterase I) (HIV-Nef associated
 DE acyl coA thioesterase) (Thioesterase II) (HTE).
 GN PTEI OR HNACTE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 CC SEQUENCE FROM N.A.
 RX MEDLINE=97445156; PubMed=9299485;
 RA Watanabe H., Shiratori T., Shoji H., Miyatake S., Okazaki Y.,
 RA Ikuta K., Sato T., Saito T.;
 RT "A novel acyl-CoA thioesterase enhances its enzymatic activity by
 RT direct binding with HIV Nef.";
 RL Biochem. Biophys. Res. Commun. 238:234-239(1997).
 RN [2]
 CC SEQUENCE FROM N.A.
 RP TISSUE=Lymphoid;

RX MEDLINE=97298085; PubMed=9153233;
 RA Liu L.X., Margottin F., LeGall S., Schwartz O., Selig L., Benarous R.,
 RA Benichou S.;
 RT "Binding of HIV-1 Nef to a novel thioesterase enzyme correlates with
 RT Nef-mediated CD4 down-regulation.";
 RL J. Biol. Chem. 272:13779-13785(1997).
 RN [3]
 CC SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=99194760; PubMed=10092594;
 RA Jones J.M., Nau K., Geraghty M.T., Erdmann R., Gould S.J.;
 RT "Identification of peroxisomal acyl-CoA thioesterases in yeast and
 RT humans.";
 RL J. Biol. Chem. 274:9216-9223(1999).
 RN [4]
 CC SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Cooby N.R.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby M.,
 RA Coulson A.G., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehaeslaih M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A.K., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 CC -!- FUNCTION: MAY PLAY A ROLE IN FATTY ACID OXIDATION RATHER THAN
 CC FORMATION OF FATTY ACIDS. MAY MEDIATE NEF-INDUCED DOWN-REGULATION
 CC OF CD4
 CC -!- CATALYTIC ACTIVITY: Palmitoyl-CoA + H(2)O = CoA + palmitate.
 CC -!- SUBUNIT: INTERACTS WITH HIV-1 NEF.
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.
 CC -!- SIMILARITY: BELONGS TO THE C/M/P THIOESTER HYDROLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF014404; AAB71655.1; -;
 CC EMBL: X86032; CAA60024.1; -;
 CC EMBL: AF124264; AAD27616.1; -;
 CC EMBL: AL008726; CAA15502.1; -;
 CC PIR: JC5644; JC5644.
 CC HSP: P23911; 1C8U.
 CC Genew: HGNC:15919; PTEI.
 CC GO: GO:0005777; C:peroxisome; TAS.
 CC GO: GO:0008778; F:acyl-CoA thioesterase II activity; TAS.
 CC GO: GO:0006960; P:antimicrobial humoral response (sensu Inver. .); TAS.
 CC GO: GO:0006629; P:lipid metabolism; TAS.
 CC InterPro: IPR003703; Acyl_COA_thio.
 CC Pfam: PF02551; Acyl_COA_thio; 2.


```
DR TIGRFAMS: TIGR00189; tesB; 1.
KW Hydroxylase; Serine esterase; Peroxisome.
FT ACT_SITE 78 78 BY SIMILARITY.
FT ACT_SITE 232 232 BY SIMILARITY.
FT SITE 317 319 MICROBODY TARGETING SIGNAL (POTENTIAL).
FT CONFLICT 291 293 LWR -> VNS (IN REF. 2).
FT CONFLICT 319 319 L -> R (IN REF. 2).
SQ SEQUENCE 319 AA; 35914 MW; 8345C6E5EABF3326 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 319;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 286 LVHGRL 291

RESULT 7
PTEI_MOUSE
ID PTEI_MOUSE STANDARD; PRT; 320 AA.
AC P58137;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxisomal acyl-coenzyme A thioester hydrolase 1 (EC 3.1.2.2)
DE (Peroxisomal long-chain acyl-coA thioesterase 1).
GN PTEI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McWain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: MAY PLAY A ROLE IN FATTY ACID OXIDATION RATHER THAN
CC -!- CATALYTIC ACTIVITY: Palmitoyl-CoA + H(2)O = CoA + palmitate.
CC -!- SUBCELLULAR LOCATION: Peroxisomal (By similarity)
CC -!- SIMILARITY: BELONGS TO THE C/M/P THIOESTER HYDROLASE FAMILY.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; BC005792; AAH05792.1; -.
DR HSSP; P23911; 1C8U,
DR MGD; MGI:2158201; Pte1.
DR GO; GO:0016291; F:acyl-CoA thioesterase activity; IDA.
```

```
DR GO; GO:0006637; P:acyl-CoA metabolism; IDA.
DR InterPro; IPR003703; Acyl_COA_thio.
DR Pfam; PF02551; Acyl_COA_thio; 2.
DR TIGRFAMS: TIGR00189; tesB; 1.
KW Hydroxylase; Serine esterase; Peroxisome.
FT ACT_SITE 79 79 BY SIMILARITY.
FT ACT_SITE 233 233 BY SIMILARITY.
FT SITE 318 320 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 320 AA; 35827 MW; 94F6AFCFEE2FA23 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 320;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 287 LVHGRL 292

RESULT 8
Y776_METJA
ID Y776_METJA STANDARD; PRT; 383 AA.
AC Q58186;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0776.
GN MJ0776.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U67522; AAB98766.1; -.
DR PIR; H64396; H64396.
DR TIGR; MJ0776; -.
DR InterPro; IPR003806; DUF201.
DR Pfam; PF02655; DUF201; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
SQ SEQUENCE 383 AA; 43948 MW; FD11A7D7E8C241AC CRC64;

Query Match 90.3%; Score 28; DB 1; Length 383;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 65 LVHGRL 70
```

```
RESULT 9
EX7L_CHLPN
ID EX7L_CHLPN STANDARD; PRT; 554 AA.
AC O926J7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR CPN1062 OR CP0787.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN SEQUENCE FROM N.A.
RP STRAIN=CNL029;
RC MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grinwood J., Davis R.W., Stephens R.S.;
RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RA Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heigelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RA pneumoniae AR39.";
RA Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RA from Japan and CWL029 from USA.";
RA Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: BIDIRECTIONALLY DEGRADABLE SINGLE-STRANDED DNA INTO LARGE
CC ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3'
CC or 3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001686; AAD19199.1; -
DR EMBL; AE002238; AAF38586.1; -
DR EMBL; AP002548; BAA99269.1; -
DR PIR; C72002; C72002.
DR PIR; C86823; C86823.
DR TIGR; CP0787; -.
DR HAMAP; MF_00378; -.
DR InterPro; IPR003753; Exonuc_VII_L.
DR Pfam; PF02601; Exonuc_VII_L.1.
DR Pfam; PF01336; tRNA_anti1; 1.
DR TIGRfams; TIGR00237; xsea; 1.

KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 554 AA; 62911 MW; 466902ABAF19F052 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 554;
Best Local Similarity 66.7%; Pred. No. 88;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKGL 6
Db 77 IIVGKL 82

RESULT 10
RECG_BACSU STANDARD; PRT; 682 AA.
ID RECG_BACSU
AC O34942;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP-dependent DNA helicase recG (EC 3.6.1.-).
GN RECG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98195738; PubMed=9534248;
RA Foulger D., Errington J.;
RA "A 28 kbp segment from the spoVM region of the Bacillus subtilis 168
RA genome.";
RA Microbiology 144:801-805(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel P.A.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Totsato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RA "The complete genome sequence of the Gram-positive bacterium Bacillus
RA subtilis.";
RA Nature 390:249-256(1997).
RL Nature 390:249-256(1997).
CC -1- FUNCTION: Critical role in recombination and DNA repair. Help
CC process Holliday junction intermediates to mature products by
CC catalyzing branch migration. Has a DNA unwinding activity
CC characteristic of a DNA helicase with a 3' to 5' polarity. RecG
CC unwind branched duplex DNA (Y-DNA) (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECG SUBFAMILY.
```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y13937; CAA74246.1; -.
DR EMBL; Z99112; CAB13460.1; -.
DR PIR; H69879; H69879; recG.
DR Subtilist; BG13399; recG.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004609; RecG.
DR InterPro; IPR004365; trna_anti.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF01336; trna_anti; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00643; recG; 1.
DR Hydrolase; Helicase; DNA repair; ATP-binding; DNA recombination;
KW DNA-binding; Complete proteome.
FT NP_BIND 284 291 ATP (POTENTIAL).
FT SITE 385 388 DEQH BOX.
SQ SEQUENCE 682 AA; 78140 MW; 5E479058D782295C CRC64;

Query Match 90.3%; Score 28; DB 1; Length 682;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 506 LMHGKL 511

RESULT 11
RECG STAAU STANDARD; PRT; 686 AA.
AC O50581;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP-dependent DNA helicase recG (EC 3.6.1.-).
GN RECG
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM4220.
RX MEDLINE=97400264; PubMed=9257758;
RA Niga T., Yoshida H., Hattori H., Nakamura S.;
RT "Cloning and sequencing of a novel gene (recG) that affects the
RT quinolone susceptibility of Staphylococcus aureus.";
RL Antimicrob. Agents Chemother. 41:1770-1774(1997).
CC -!- FUNCTION: Critical role in recombination and DNA repair. Help
CC process Holliday junction intermediates to mature products by
CC catalyzing branch migration. Has a DNA unwinding activity
CC characteristic of a DNA helicase with a 3' to 5' polarity. RecG
CC unwind branched duplex DNA (Y-DNA) (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECG SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AB000439; BAA24572.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004609; RecG.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00643; recG; 1.
DR Hydrolase; Helicase; DNA repair; ATP-binding; DNA recombination;
KW DNA-binding.
FT NP_BIND 292 299 ATP (POTENTIAL).
FT SITE 392 395 DEQH BOX.
SQ SEQUENCE 686 AA; 78343 MW; D3405F218D45432B CRC64;

Query Match 90.3%; Score 28; DB 1; Length 686;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 513 LLHGKL 518

RESULT 12
NTT7_BOVIN STANDARD; PRT; 729 AA.
AC Q9X559;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Orphan sodium- and chloride-dependent neurotransmitter transporter
DE NTT73 (Orphan transporter v7-3).
GN NTT73.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=99148812; PubMed=10025945;
RA Sakata K., Shimada S., Yamashita T., Inoue K., Tohyama M.;
RT "Cloning of a bovine orphan transporter and its short splicing
RT variant.";
RL FEBS Lett. 443:267-270(1999).
CC -!- FUNCTION: NOT YET KNOWN, ORPHAN TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNF)
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB020854; BAA77223.1; -.
DR InterPro; IPR000175; Na/ntran_symport.
DR Pfam; PF00209; SNF; 1.
DR PRINTS; PR00176; NANEUSMPORT.
DR Prodom; PD000448; Na/ntran_symport; 2.
DR PROSITE; PS00610; NA_NEUROTRAN_SYM_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYM_2; FALSE_NEG.
DR PROSITE; PS0267; NA_NEUROTRAN_SYM_3; 1.
KW Neurotransmitter transporter; Transport; Transmembrane; Glycoprotein;
KW Symport.
FT DOMAIN 1 69 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 70 90 1 (POTENTIAL).
FT TRANSMEM 98 117 2 (POTENTIAL).

```

```

FT TRANSMEM 142 162 3 (POTENTIAL).
FT DOMAIN 163 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 244 4 (POTENTIAL).
FT TRANSMEM 253 270 5 (POTENTIAL).
FT TRANSMEM 306 323 6 (POTENTIAL).
FT TRANSMEM 335 356 7 (POTENTIAL).
FT DOMAIN 357 452 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 453 472 8 (POTENTIAL).
FT TRANSMEM 496 514 9 (POTENTIAL).
FT TRANSMEM 530 550 10 (POTENTIAL).
FT TRANSMEM 571 592 11 (POTENTIAL).
FT TRANSMEM 620 642 12 (POTENTIAL).
FT DOMAIN 643 729 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 729 AA; 81681 MW; 332FDB3349C196A9 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 729;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
   I:|:|:|
Db 676 LIHGKI 681

RESULT 13
NTT7_RAT
ID NTT7_RAT STANDARD; PRT; 729 AA.
AC Q08469; Q63838;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Orphan sodium- and chloride-dependent neurotransmitter transporter
DE NTT73 (Orphan transporter v7-3).
GN NTT73.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93180651; PubMed=1363329;
RA Uhl G.R., Kitayama S., Gregor P., Nanthakumar E., Persico A.M.,
RA Shimada S.;
RT "Neurotransmitter transporter family cDNAs in a rat midbrain library:
RT 'Orphan transporters' suggest sizable structural variations. ";
RL Brain Res. Mol. Brain Res. 16:353-359(1992).
CC -!- FUNCTION: NOT YET KNOWN, ORPHAN TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNF)
CC family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L22022; AAA41729.1; -
DR EMBL; S56968; AAB25532.1; -
DR InterPro; IPR000175; Na/ntran_symport.
DR Pfam; PF00209; SNF; 1.
DR PRINTS; PR00176; NAMEUSMPORT.
DR PRODOM; PD000448; Na/ntran_symport; 2.
DR PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYMP_2; FALSE_NEG.
DR PROSITE; PS50267; NA_NEUROTRAN_SYMP_3; 1.

```

```

KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport.
FT DOMAIN 1 69 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 70 90 1 (POTENTIAL).
FT TRANSMEM 98 117 2 (POTENTIAL).
FT TRANSMEM 142 162 3 (POTENTIAL).
FT DOMAIN 163 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 244 4 (POTENTIAL).
FT TRANSMEM 253 270 5 (POTENTIAL).
FT TRANSMEM 306 323 6 (POTENTIAL).
FT TRANSMEM 335 356 7 (POTENTIAL).
FT DOMAIN 357 452 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 453 472 8 (POTENTIAL).
FT TRANSMEM 496 514 9 (POTENTIAL).
FT TRANSMEM 530 550 10 (POTENTIAL).
FT TRANSMEM 571 592 11 (POTENTIAL).
FT TRANSMEM 620 642 12 (POTENTIAL).
FT DOMAIN 643 729 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1 3 MPK -> IP (IN REF. 1; AAB25532).
SQ SEQUENCE 729 AA; 81596 MW; ECDDC6403741B53E CRC64;

Query Match 90.3%; Score 28; DB 1; Length 729;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
   I:|:|:|
Db 676 LIHGKI 681

RESULT 14
NTT7_HUMAN
ID NTT7_HUMAN STANDARD; PRT; 730 AA.
AC Q9H2J7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Orphan sodium- and chloride-dependent neurotransmitter transporter
DE NTT73 (Orphan transporter v7-3).
GN NTT73.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RX MEDLINE=20564208; PubMed=11112352;
RA Farmer M.K., Robbins M.J., Medhurst A.D., Campbell D.A., Ellington K.,
RA Duckworth M., Brown A.M., Middlemiss D.N., Price G.W., Pangalos M.N.;
RT "Cloning and characterization of human NTT5 and v7-3: two orphan
RT transporters of the Na(+)/Cl(-)-dependent neurotransmitter
RT transporter gene family. ";
RL Genomics 70:241-252(2000).
CC -!- FUNCTION: NOT YET KNOWN, ORPHAN TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNF)
CC family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF265577; AAG41361.1; -
DR GO; GO:0006836; P:neurotransmitter transport; NAS.
DR InterPro; IPR000175; Na/ntran_symport.

```

```

DR Pfam; PF00209; SNF; 1.
DR PRINTS; PR00176; NANEUSMPORT.
DR ProDom; PD000448; Na/utran_symp; 2.
DR PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYMP_2; FALSE_NEG.
DR PROSITE; PS00267; NA_NEUROTRAN_SYMP_3; 1.
DR Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport.
FT DOMAIN 1 69 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 70 90 1 (POTENTIAL).
FT TRANSMEM 98 117 2 (POTENTIAL).
FT TRANSMEM 142 162 3 (POTENTIAL).
FT DOMAIN 163 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 244 4 (POTENTIAL).
FT TRANSMEM 253 270 5 (POTENTIAL).
FT TRANSMEM 306 323 6 (POTENTIAL).
FT TRANSMEM 335 356 7 (POTENTIAL).
FT DOMAIN 357 452 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 453 472 8 (POTENTIAL).
FT TRANSMEM 496 514 9 (POTENTIAL).
FT TRANSMEM 530 550 10 (POTENTIAL).
FT TRANSMEM 571 592 11 (POTENTIAL).
FT TRANSMEM 620 642 12 (POTENTIAL).
FT DOMAIN 643 730 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 730 AA; 81836 MW; 45963118E06CFFE6 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 730;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 676 LIHGKI 681

RESULT 15
AKRL_YEAST STANDARD; PRT; 764 AA.
AC P39010;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin repeat-containing protein AKRL
GN AKRL OR YDR264C OR D9954.9 OR YD9230B.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
R Bender A., Peterson J.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Willson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 674-764 FROM N.A.
RC STRAIN=S288c / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 6 ANK repeats.

```

```

CC -!- SIMILARITY: Contains 1 DHHC-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L31407; AAC41676.1; -
CC EMBL; U51030; AAB64454.1; -
CC EMBL; Z70202; CAA94103.1; -
CC EMBL; Z68290; CAA92582.1; -
CC PIR; S48521; S48521.
CC HSSP; P80144; 2MYO.
CC SGD; S0002672; AKRL.
CC GO; GO:0016020; C:membrane; IDA.
CC GO; GO:0016409; F:palmitoyltransferase activity; IDA.
CC GO; GO:0006897; P:endocytosis; IMP.
CC GO; GO:0018318; P:protein amino acid palmitoylation; IDA.
CC GO; GO:0000750; P:signal transduction during conjugation with. . .; IGI.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001594; Znf_DHHC.
CC Pfam; PF00023; ank; 5.
CC Pfam; PF01529; zf-DHHC; 1.
CC SMART; SM00248; ANK; 6.
CC PROSITE; PS50088; ANK_REPEAT; 4.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS50216; ZF_DHHC; 1.
CC ANK repeat; Repeat; Zinc-finger.
FT REPEAT 72 102 ANK 1.
FT REPEAT 108 137 ANK 2.
FT REPEAT 142 171 ANK 3.
FT REPEAT 175 204 ANK 4.
FT REPEAT 213 242 ANK 5.
FT REPEAT 246 275 ANK 6.
FT ZN_FING 470 520 DHHC-TYPE.
SQ SEQUENCE 764 AA; 85840 MW; 9C9759F0140CE3F6 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 764;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 92 MIHGKL 97

RESULT 16
MTP_BOVIN STANDARD; PRT; 887 AA.
AC P55156;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Microsomal triglyceride transfer protein, large subunit precursor
DE (Fragment).
GN MTP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=93368660; PubMed=8361539;
RA Sharp D., Blinderman L., Combs K.A., Kienzle B., Ricci B.,
RA Wager-Smith K., Gil C.M., Turk C.W., Bouma M.-E., Rader D.J.,
RA Aggerberk L.P., Gregg R.E., Gordon D.A., Wetterau J.R.;
RT "Cloning and gene defects in microsomal triglyceride transfer protein
RT associated with abetalipoproteinemia.";

```

RL Nature 365:65-69(1993).
CC -!- FUNCTION: CATALYZES THE TRANSPORT OF TRIGLYCERIDE, CHOLESTERYL
CC ESTER, AND PHOSPHOLIPID BETWEEN PHOSPHOLIPID SURFACES. REQUIRED
CC FOR THE SECRETION OF PLASMA LIPOPROTEINS THAT CONTAIN
CC APOLIPOPROTEIN B.
CC
CC -!- SUBUNIT: HETERODIMER COMPOSED OF MTP AND OF PROTEIN DISULFIDE
CC ISOMERASE (PDI).
CC
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC
CC -!- SIMILARITY: TO VITELLOGENINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78567; CAA55310.1; -
CC PIR; A46764; A46764.
CC InterPro; IPR001747; Lipid_transprt_N.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC SMART; SM00638; LPD_N; 1.
CC Endoplasmic reticulum; Lipid-binding; Transport; Signal.
FT NON_TER 1 1
FT SIGNAL <1 11 POTENTIAL.
FT CHAIN 12 887 MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN,
FT LARGE SUBUNIT.
SQ SEQUENCE 887 AA; 99031 MW; 45CD454541054BBF CRC64;

Query Match 90.3%; Score 28; DB 1; Length 887;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 120 LIHGKI 125
:::||||

RESULT 17
MFD_BORBU
ID MFD_BORBU STANDARD; PRT; 1125 AA.
AC O51568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription-repair coupling factor (TRCF).
CN MFD OR BB0623.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.N., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wathley L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi".
RL Nature 390:580-586(1997).
CC -!- FUNCTION: NECESSARY FOR STRAND-SPECIFIC REPAIR. A LESION IN THE
CC TEMPLATE STRAND BLOCKS THE RNA POLYMERASE COMPLEX (RNAP). THE
CC RNAP-DNA-RNA COMPLEX IS SPECIFICALLY RECOGNIZED BY TRCF WHICH
CC RELEASES RNAP AND THE TRUNCATED TRANSCRIPT; THE TRCF MAY REPLACE
CC RNAP AT THE LESION SITE AND THEN RECRUIT THE UVRA/B/C REPAIR
CC SYSTEM (BY SIMILARITY).
CC
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE UVRB FAMILY.

CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HELICASE
CC FAMILY. RECG SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001163; AAC66973.1; -
CC PIR; F70177; F70177.
CC TIGR; BB0623; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004576; Mfd.
CC InterPro; IPR005118; TRCF.
CC Pfam; PF02559; Card_TRCF; 1.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF03461; TRCF; 1.
CC SMART; SM00487; DEXdc; 1.
CC SMART; SM00490; HELICC; 1.
CC TIGRFAMS; TIGR00580; mfd; 1.
CC Helicase; DNA repair; ATP-binding; DNA-binding; Complete proteome.
KW NP_BIND 610 617 ATP (POTENTIAL).
FT SITE 711 714 DEEQ BOX.
FT SEQUENCE 1125 AA; 130729 MW; 116FBF8DE9024539 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 1125;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 822 IIHGKL 827
:::||||

RESULT 18
CYAL_DROME
ID CYAL_DROME STANDARD; PRT; 2248 AA.
AC P32870;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ca(2+)/calmodulin-responsive adenylyl cyclase (EC 4.6.1.1) (ATP
DE pyrophosphate-lyase) (Rutabaga protein).
CN RUT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; TISSUP=Head;
RX MEDLINE=92154664; PubMed=1739965;
RA Levin L.R., Han P.-L., Hwang P.M., Feinstein P.G., Davis R.L.,
RA Reed R.R.;
RT "The Drosophila learning and memory gene rutabaga encodes a
RT Ca2+/calmodulin-responsive adenylyl cyclase".
RL Cell 68:479-489(1992).
CC -!- FUNCTION: This is a membrane-bound, calmodulin-sensitive adenylyl
CC cyclase. Inactivation of this cyclase leads to a learning and
CC memory defect.
CC
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -!- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
CC -!- ENZYME REGULATION: Activated by calcium/calmodulin and G protein.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.
CC
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC
CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl

```

CC      cyclase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to licenses@isb-sib.ch).
CC      -----
CC      EMBL: M81887; AAA28844.1; -
CC      DR      PIR: D42088; D42088.
CC      DR      HSSP: P19734; LAWK.
CC      DR      FlyBase: FBgn0003301; rut.
CC      DR      GO: GO:0019333; P:cAMP-mediated signaling; NAS.
CC      DR      GO: GO:0007625; P:grooming behavior; NAS.
CC      DR      GO: GO:0007591; P:molting cycle (sensu Insecta); IGI.
CC      DR      GO: GO:0008355; P:olfactory learning; NAS.
CC      DR      GO: GO:0043473; P:response to ethanol (sensu Insecta); NAS.
CC      DR      InterPro: IPR001094; G_cyclase.
CC      DR      Pfam: PF00211; guanylate_cyc; 2.
CC      DR      SMART: SM00044; CYCC; 2.
CC      DR      PROSITE: PS00452; GUANYLATE_CYCLASES.1; 2.
CC      DR      PROSITE: PS00125; GUANYLATE_CYCLASES.2; 2.
CC      DR      Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
KW      Metal-binding; Magnesium.
KW      DOMAIN 1 41
FT      TRANSMEM 42 60      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 65 84      POTENTIAL.
FT      TRANSMEM 101 115    POTENTIAL.
FT      TRANSMEM 122 142    POTENTIAL.
FT      TRANSMEM 152 174    POTENTIAL.
FT      TRANSMEM 186 206    POTENTIAL.
FT      DOMAIN 207 705      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 706 726    POTENTIAL.
FT      TRANSMEM 730 750    POTENTIAL.
FT      TRANSMEM 770 791    POTENTIAL.
FT      DOMAIN 792 813      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 814 834    POTENTIAL.
FT      TRANSMEM 842 867    POTENTIAL.
FT      TRANSMEM 868 888    POTENTIAL.
FT      DOMAIN 889 2248     CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 318 334    CATALYTIC (POTENTIAL).
FT      DOMAIN 1013 1029    CATALYTIC (POTENTIAL).
FT      DOMAIN 515 530      GLY-RICH.
FT      DOMAIN 569 602      GLY-RICH.
FT      DOMAIN 1278 1297    GLN-RICH.
FT      DOMAIN 1767 1810    GLY/SER-RICH.
FT      DOMAIN 2025 2040    ASP/GLU-RICH (ACIDIC).
FT      DOMAIN 2200 2241    GLN-RICH.
FT      METAL 280 280      MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT      METAL 281 281      MAGNESIUM 2 (VIA CARBONYL OXYGEN) (BY
FT      METAL 281 281      SIMILARITY).
FT      METAL 324 324      MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT      CARBOHYD 800 800    N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 807 807    N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      MUTAGEN 1026 1026  G>R: ABOLISHES CATALYTIC ACTIVITY.
SQ      SEQUENCE 2248 AA; 248899 MW; E459C718BE018868 CRC64;

Query Match          90.3%; Score 28; DB 1; Length 2248;
Best Local Similarity 83.3%; Pred. No. 3.8e-02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 LVHGKL 6
      111111
Db      893 LVHGRL 898

RESULT 19
RHEB_HUMAN
ID      RHEB_HUMAN          STANDARD;          PRT;          184 AA.
AC      Q15382; Q99444;
DT      30-MAY-2000 (Rel. 39, Created)

```

```

DT      30-MAY-2000 (Rel. 39, Last sequence update)
DE      GTP-binding protein Rheb (Ras homolog enriched in brain 2).
GN      RHEB2 OR RHEB.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Psoriatic skin;
RX      MEDLINE=96128233; PubMed=8543055;
RA      Gromov P.S., Madsen P., Tomerup N., Celis J.E.;
RT      "A novel approach for expression cloning of small GTPases:
RT      identification, tissue distribution and chromosome mapping of the
RT      human homolog of rheb."
RL      FEBS Lett. 377:221-226(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RX      MEDLINE=96299667; PubMed=8661031;
RA      Mizuki N., Kimura M., Ohno S., Miyata S., Sato M., Ando H.,
RA      Ishihara M., Goto K., Watanabe S., Yamazaki M., Ono A., Taguchi S.,
RA      Okumura K., Nogami M., Taguchi H., Ando A., Inoko H.;
RT      "Isolation of cDNA and genomic clones of a human Ras-related GTP-
RT      binding protein gene and its chromosomal localization to the long arm
RT      of chromosome 7, 7q36."
RL      Genomics 34:114-118(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      Weidenmuller U., Tur M.K., Tawadros S., Engert A., Barth S.;
RT      "Detection of membrane-associated proteins using serum of mice
RT      immunized with membrane fractions of breast carcinoma cells."
RL      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RX      Abu-Threideh J., Le T.;
RL      Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN      [5]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
RL      Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN      [6]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Muscle;
RX      MEDLINE=22388257; PubMed=12477932;
RA      Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA      Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA      Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length
RT      human and mouse cDNA sequences."
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC      -!- FUNCTION: NOT KNOWN: BINDS GTP AND EXHIBITS INTRINSIC GTPASE
CC      -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGHEST LEVELS OBSERVED IN
CC      SKELETAL AND CARDIAC MUSCLE.
CC      -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHEB FAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z29677; CAA82774.1; -
 DR EMBL; D78132; BAA1121.1; -
 DR EMBL; AF148645; AAF73125.1; -
 DR EMBL; AC005996; AAD15548.1; -
 DR EMBL; AF493921; AAM12635.1; -
 DR EMBL; BC016155; AAH16155.1; -
 DR PIR; S68419; S41960.
 DR HSP; P10113; IGDA.
 DR Gene; HGNC:10011; RHEB2.
 DR MIM; 601293; -
 DR GO; GO:0016020; C:membrane; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR InterPro; IPR003577; GTPase_Ras.
 DR InterPro; IPR001806; Ras_trnsmg.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00071; ras; 1.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR SMART; SM00173; RAS; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 DR GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 13 20 GTP (BY SIMILARITY).
 FT NP_BIND 60 64 GTP (BY SIMILARITY).
 FT NP_BIND 119 122 GTP (BY SIMILARITY).
 FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).
 FT LIPID 181 FARNESYL (BY SIMILARITY).
 FT CONFLICT 118 118 G -> W (IN REF. 2).
 SQ SEQUENCE 184 AA; 20497 MW; 8F4C8080BDF928FD CRC64;

Query Match 87.1%; Score 27; DB 1; Length 184;
 Best Local Similarity 66.7%; Pred. No. 48;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 98 VIHGKL 103

RESULT 20
 RHEB_MOUSE
 ID RHEB_MOUSE STANDARD; PRT; 184 AA.
 AC Q92LJ2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE GTP-binding protein RHEB.
 GN RHEB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Mailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: NOT KNOWN; BINDS GTP AND EXHIBITS INTRINSIC GTPASE
 CC ACTIVITY (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHEB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; BC012273; AAH12273.1; -
 DR HSP; P05713; 3RAB.
 DR MGD; MGI:97912; Rheb.
 DR GO; GO:0005681; C:spliceosome complex; IDA.
 DR InterPro; IPR003577; GTPase_Ras.
 DR InterPro; IPR001806; Ras_trnsmg.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00071; ras; 1.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR SMART; SM00173; RAS; 1.
 DR TIGRFAMs; TIGR00231; small_GTP; 1.
 DR GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 13 20 GTP (BY SIMILARITY).
 FT NP_BIND 60 64 GTP (BY SIMILARITY).
 FT NP_BIND 119 122 GTP (BY SIMILARITY).
 FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).
 FT LIPID 181 FARNESYL (BY SIMILARITY).
 SQ SEQUENCE 184 AA; 20451 MW; 7EC58080BDF92DFB CRC64;

Query Match 87.1%; Score 27; DB 1; Length 184;
 Best Local Similarity 66.7%; Pred. No. 48;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 :|||||
 Db 98 VIHGKL 103

RESULT 21
 RHEB_RAT
 ID RHEB_RAT STANDARD; PRT; 184 AA.
 AC Q62639;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GTP-binding protein Rheb.
 GN RHEB.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; Tissue=Hippocampus;
 RX MEDLINE=94266828; PubMed=8206940;
 RA Yamagata K., Sanders L.K., Kaufmann W.E., Yee W., Barnes C.A.,
 RA Nathans D., Worley P.F.;
 RT "Rheb, a growth factor- and synaptic activity-regulated gene, encodes
 RT a novel Ras-related protein".
 RL J. Biol. Chem. 269:16333-16339(1994).
 CC -!- FUNCTION: NOT KNOWN; BINDS GTP AND EXHIBITS INTRINSIC GTPASE
 CC ACTIVITY.

CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN NORMAL ADULT
CC CORTEX AS WELL AS A NUMBER OF PERIPHERAL TISSUES, INCLUDING LUNG
CC AND INTESTINE.
CC
CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHEB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U08227; AAA21380.1; -
DR PIR; I55401; I55401.
DR HSSP; P10113; LGUA.
DR InterPro; IPR003577; GTPase_Ras.
DR InterPro; IPR001806; Ras_transf_rmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF000071; ras; 1.
DR SMART; SM00173; Ras; 1.
DR TIGREMS; TIGR00231; small_GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 13 20 GTP (BY SIMILARITY).
FT NP_BIND 60 64 GTP (BY SIMILARITY).
FT NP_BIND 119 122 GTP (BY SIMILARITY).
FT DOMAIN 35 43 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 181 181 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 184 AA; 20479 MW; 8F4C8080BDF92DE9 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 184;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 98 VHGKL 103

RESULT 22
HIS7_THEME
ID HIS7_THEME STANDARD; PRT; 195 AA.
AC G9X0C9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) (IGPD).
GN HISB OR TM1039.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Otterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
CC -1- CATALYTIC ACTIVITY: D-erythro-1-(imidazol-4-yl)glycerol 3-
CC phosphate + 3-(imidazol-4-yl)-2-oxopropyl phosphate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; sixth step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the imidazoleglycerol-phosphate dehydratase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001764; AAD36116.1; -
DR PIR; F72304; F72304.
DR TIGR; TM1039; -
DR HAMAP; MF_00076; -; 1.
DR InterPro; IPR000807; IGPD.
DR Pfam; PF00475; IGPD; 1.
DR ProDom; PD002282; IGPD; 1.
DR PROSITE; PS00954; IGP_DEHYDRATASE_1; FALSE_NEG.
DR PROSITE; PS00955; IGP_DEHYDRATASE_2; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
SQ SEQUENCE 195 AA; 21997 MW; 719E6B9A6F610E73 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGKL 6
DB 28 VHGKL 32

RESULT 23
TCPL_HUMAN
ID TCPL_HUMAN STANDARD; PRT; 215 AA.
AC Q8TRD4; Q96LN5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE T-complex protein 10A homolog 2 (T-complex protein 10A-2) (TCP10A-2)
DE (TCP10-like).
DE TCPI0L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ge H.P., Yu L., Jin L., Fan Y.X., Yang Y.M., Zhao S.Y.;
RT "Cloning of a new human cDNA homologous to human T-complex protein 10A
RT mRNA."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA Kawakami B., Nagai K., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Pooley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Vaspo M.-L.;

```

RT  "The DNA sequence of human chromosome 21.";
RL  Nature 405:311-319(2000).
RN  [4]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=22388257; PubMed=12477932;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Brownstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Villalon D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA  Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA  Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length
RL  human and mouse cDNA sequences.";
RN  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP  CC -1- SIMILARITY: BELONGS TO THE TCP10 FAMILY.
RX  CC -----
RA  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
RA  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA  CC the European Bioinformatics Institute. There are no restrictions on its
RA  CC use by non-profit institutions as long as its content is in no way
RA  CC modified and this statement is not removed. Usage by and for commercial
RA  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RA  CC or send an email to license@isb-sib.ch).
RA  CC -----
DR  EMBL; AF115967; AAM18531.1; .
DR  EMBL; AK058078; BAB71653.1; .
DR  EMBL; AP001714; -; NOT_ANNOTATED_CDS.
DR  EMBL; BC033855; AAH33855.1; .
DR  Genew; HGNC:11657; TCP10L.
KW  Coiled coil. 69 110 COILED COIL (POTENTIAL).
FT  DOMAIN 145 145 H -> R (IN REF. 3).
FT  CONFLICT 145 145
SQ  SEQUENCE 215 AA; 23863 MW; CF3E74AB520F5FE CRC64;

Query Match 87.1%; Score 27; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VHGL 6
Db 71 VHGL 75

RESULT 24
SRY_HORSE STANDARD; PRT; 223 AA.
AC P36389; 018729; P79354;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Sex-determining region Y protein (Testis-determining factor).
GN SRY OR TDF.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Thoroughbred; TISSUE=Testis;
RA MEDLINE=99151238; PubMed=10027176;
RA Hasegawa T., Ishida M., Harigaya T., Sato F., Ishida N., Mukoyama H.;

"Linear SRY transcript in equine testis.";
J. Vet. Med. Sci. 61:97-100(1999).
[2]
RP SEQUENCE OF 54-108 FROM N.A.
RX MEDLINE=97451094; PubMed=9306070;
RA Meyers-Wallen V.N., Hurlgen J., Schlafer D., Tulleners E.,
RA Cleland W.R., Ruth G.R., Acland G.M.;
RT "Sry-negative XX true hermaphroditism in a Pasa Fino horse.";
RL Equine Vet. J. 29:404-408(1997).
RN [3]
RP SEQUENCE OF 62-116 FROM N.A.
RX MEDLINE=94214723; PubMed=81622230;
RA Griffiths R., Tiwari B.;
RT "Primers for the differential amplification of the sex-determining
RA region Y gene in a range of mammal species.";
RL Mol. Ecol. 2:405-406(1993).
RP CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC
RX SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE
RA SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL
CC INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 HMG box domain.
CC -----
RA  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
RA  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA  CC the European Bioinformatics Institute. There are no restrictions on its
RA  CC use by non-profit institutions as long as its content is in no way
RA  CC modified and this statement is not removed. Usage by and for commercial
RA  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
RA  CC or send an email to license@isb-sib.ch).
RA  CC -----
DR  EMBL; AB004572; BAA22428.1; .
DR  EMBL; U66067; AAC10937.1; .
DR  EMBL; Z26908; CAA81534.1; .
DR  HSPSP; Q05066; IHRV.
DR  InterPro; IPR000910; HMG_12_box.
DR  Pfam; PF00505; HMG_box; 1.
DR  SMART; SM00398; HMG; 1.
DR  PROSITE; PS50118; HMG_BOX_2; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Sexual differentiation.
FT DNA_BIND 54 122 HMG BOX.
SQ SEQUENCE 223 AA; 25353 MW; 3A5C34F8B272D0ED CRC64;

Query Match 87.1%; Score 27; DB 1; Length 223;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGL 6
Db 162 LVHGL 167

RESULT 25
SFS CAUCR STANDARD; PRT; 239 AA.
ID SFS CAUCR
AC Q9A426;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sugar fermentation stimulation protein homolog.
GN SFS OR CC2675.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RA MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

```

```

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.:
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- SIMILARITY: BELONGS TO THE SFSA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE005935; AAK24642.1; -.
DR PIR: F87580; F87580.
DR TIGR: CF2675; -.
DR HAMAP: MF_00095; -.
DR InterPro: IPR005224; Sfsa.
DR Pfam: PF03749; Sfsa; 1.
DR TIGRFAMs: TIGR00230; sfsa; 1.
DR Complete proteome.
KW SEQUENCE 239 AA; 26240 MW; AC1FDF57361B9BB1 CRC64;
SQ
Query Match 87.1%; Score 27; DB 1; Length 239;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 7 LIHGRL 12

RESULT 26
NAGB_BACHD
ID NAGB_BACHD STANDARD; PRT; 246 AA.
AC Q9KF08;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucosamine-6-phosphate deaminase (EC 3.5.99.6) (Glucosamine-6-
DE phosphate isomerase) (GNPDA) (GlcNGP deaminase).
GN NAGB OR BH0420.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Catalyzes the reversible isomerization-deamination of
CC glucosamine 6-phosphate (GlcNGP) to form fructose 6-phosphate
CC (Fru6P) and ammonium ion (By similarity).
CC -!- CATALYTIC ACTIVITY: D-glucosamine 6-phosphate + H(2)O = D-fructose
CC 6-phosphate + NH(3).
CC -!- PATHWAY: N-acetylglucosamine utilization.
CC -!- SIMILARITY: Belongs to the glucosamine/galactosamine-6-phosphate
CC isomerase family. Nagb subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001508; BAB04139.1; -.
DR HSSP: P09375; IDEA.
DR HAMAP: MF_01241; -.
DR InterPro: IPR006148; Gluc_gal_isom.
DR Pfam: PF01182; Glucosamine_iso; 1.
DR TIGRFAMs: TIGR00502; nagB; 1.
DR PROSITE: PS01161; GLC_GALNAC_ISOMERASE; 1.
KW Carbohydrate metabolism; Hydrolase; Complete proteome.
FT ACT_SITE 67 GENERAL BASE CATALYZING THE GLCNGP
FT ACT_SITE 67 ENOLIZATION STEP (BY SIMILARITY).
FT ACT_SITE 136 PART OF THE CATALYTIC TRIAD (BY
FT ACT_SITE 136 SIMILARITY).
FT ACT_SITE 138 GENERAL BASE INVOLVED IN THE CATALYSIS OF
FT ACT_SITE 143 THE RING-OPENING STEP OF GLCNGP; PART OF
FT ACT_SITE 143 THE CATALYTIC TRIAD (BY SIMILARITY).
FT ACT_SITE 143 PART OF THE CATALYTIC TRIAD (BY
FT ACT_SITE 143 SIMILARITY).
SQ SEQUENCE 246 AA; 27520 MW; 640E25631C9EC6E6 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 246;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 209 LIHGKV 214

RESULT 27
CODV_BACSU
ID CODV_BACSU STANDARD; PRT; 304 AA.
AC P39776;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable integrase/recombinase codv.
GN CODV.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX Slack F.J., Serror P., Joyce E., Sonenshein A.L.;
RA "A gene required for nutritional repression of the Bacillus subtilis
RA dipeptide permease operon.";
RL Mol. Microbiol. 15:689-702(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borrill S., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

```

RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vennier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-85 FROM N.A.
RC STRAIN=168;
RA Foulger D., Errington J.;
RT "Cloning and sequencing 7.5 Kbp of DNA from *Bacillus subtilis* upstream
RT of the *codV* gene";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13634; AB033369.1; -;
DR EMBL; 299112; CAB13487.1; -;
DR EMBL; AJ000975; CAA04424.1; -;
DR PIR; G59601; G59601.
DR HSSP; P21891; 1AOP.
DR Subtilist; BG10965; *codV*.
DR InterPro; IPR004107; Phage_integr_N.
DR Pfam; PF02899; Phage_integr_N; 1.
DR Pfam; PF00589; Phage_integrase; 1.
KW DNA recombination; DNA integration; Complete proteome.
FT ACT_SITE 281
FT TRANSIENT COVALENT LINKAGE TO DNA DURING
FT STRAND CLEAVAGE AND REJOINING (BY
FT SIMILARITY).
SQ SEQUENCE 304 AA; 35342 MW; B0778BDBDA065A053 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVHGK 5
DB 169 LVHGK 173
RESULT 28
ID BXDL_ORYSA STANDARD; PRT; 310 AA.
AC Q9AWM9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Brix domain containing protein 1 homolog.
GN P0504D03.20.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0504D03.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: Contains 1 Brix domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP002970; BAB32971.1; -;
DR Gramene; Q9AWM9; -;
DR InterPro; IPR007109; Brix.
DR Pfam; PF04427; Brix; 1.
KW Hypothetical protein; Nuclear protein.
FT DOMAIN 29 248
FT BRIX.
SQ SEQUENCE 310 AA; 35214 MW; 0E50D69F23F79DD0 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VHGLK 6
DB 262 VHGLK 266
RESULT 29
ID RLUD_BUCAI STANDARD; PRT; 312 AA.
AC P57481;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylylase)
GN RLUD OR BU401.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: RESPONSIBLE FOR SYNTHESIS OF PSEUDOURIDINE FROM URACIL
CC AT TWO POSITIONS IN 23S RIBOSOMAL RNA (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE RLUD FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BAB13104.1; -;
DR InterPro; IPR006225; Pseud_RLud.
DR InterPro; IPR006145; Pseudou_synth.
DR InterPro; IPR006224; PSI_RLUD.

DR	InterPro: IPR002942; S4.
DR	Pfam: PF00849; PseudoU_synth_2; 1.
DR	Pfam: PF01479; S4; 1.
DR	ProDom: PD001819; PSI_RLU; 1.
DR	SMART: SM00363; S4; 1.
DR	TIGRFAMS: TIGR00005; rluD_subfam; 1.
DR	PROSITE: PS01129; PSI_RLU; 1.
DR	PROSITE: PS0889; S4; 1.
KW	Lysase; RNA-binding; Complete proteome.
FT	DOMAIN 18 87 S4 RNA-BINDING.
SQ	SEQUENCE 312 AA; 35840 MW; 6AC32A7F9260D890 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 312;	
Best Local Similarity 66.7%; Pred.No. 82;	
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps	
Oy	1 LVHGKL 6
Dd	: :
	174 IVHGKM 179
RESULT 30	
MCH_ARCFU	STANDARD; PRT; 316 AA.
ID	MCH_ARCFU
AC	O28344;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	N(5).N(10)-methenyltetrahydromethanopterin cyclohydrolyase
DE	(EC 3.5.4.27) (Methenyl-H4MPT cyclohydrolyase).
GN	MCH OR AF1935.
OS	Archaeoglobus fulgidus.
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC	Archaeoglobaceae; Archaeoglobus.
OX	NCBI_TaxID=2234;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA	MEDLINE=98049343; PubMed=9389475;
RA	Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA	Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA	Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,
RA	Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA	Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA	Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA	Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA	Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA	Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA	Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA	Venter J.C.;
RT	The complete genome sequence of the hyperthermophilic, sulphate-
RT	reducing archaeon Archaeoglobus fulgidus.";
RL	Nature 390:364-370(1997).
RL	[2]
RP	SEQUENCE OF 1-40, AND CHARACTERIZATION.
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA	MEDLINE=93243882; PubMed=8481088;
RA	Klein A.R., Breitung J., Linder D., Stetter K.O., Thauer R.K.;
RT	"N5,N10-methenyltetrahydromethanopterin cyclohydrolyase from the
RT	extremely thermophilic sulfate reducing Archaeoglobus fulgidus:
RT	comparison of its properties with those of the cyclohydrolyase from
RT	the extremely thermophilic Methanopyrus kandleri.";
RL	Arch. Microbiol. 159:213-219(1993).
CC	-1- FUNCTION: Reversible interconversion of N(5)-formyl-H(4)MPT to
CC	methenyl-H(4)MPT(+).
CC	-1- CATALYTIC ACTIVITY: 5,10-methenyl-5,6,7,8-tetrahydromethanopterin
CC	+ H2O = N(5)-formyl-5,6,7,8-tetrahydromethanopterin.
CC	-1- PATHWAY: Methanogenesis (reduction of carbon dioxide to methane);
CC	third step.
CC	-1- SUBCELLULAR LOCATION: Homodimer.
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.
CC	-1- MISCELLANEOUS: OPTIMALLY ACTIVE AT 85 DEGREES CELSIUS. REQUIRES
CC	A HIGH SALT CONCENTRATION FOR OPTIMAL THERMOSTABILITY AT 90

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RN Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-30.
RC STRAIN-168 / JH642;
RX MEDLINE-96345629; PubMed-8755892;
RA Graumann P., Schroeder K., Schmid R., Marahiel M.A.;
RT "Cold shock stress-induced proteins in *Bacillus subtilis*.";
RL J. Bacteriol. 178:4611-4619(1996).
RN [4]
RP CHARACTERIZATION
RX MEDLINE-20261518; PubMed-10799476;
RA Fillinger S., Boschi-Muller S., Azza S., Dervyn E., Branlant G.,
RA Aymerich S.;
RT "Two glyceraldehyde-3-phosphate dehydrogenases with opposite
RT physiological roles in a nonphotosynthetic bacterium.";
RL J. Biol. Chem. 275:14031-14037(2000).
CC -!- FUNCTION: More active in catabolism.
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC -!- PATHWAY: Second phase of glycolysis; first step.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13011; CAA31434.1; -;
DR EMBL; 299121; CAB15399.1; -;
DR PIR; S02754; DERSEG.
DR HSSP; P00362; IGD1.
DR Subtilist; BG10827; gapA.
DR InterPro; IPR000173; GAP_dhdrogenase.
DR InterPro; IPR006424; GAPDH-I.
DR Pfam; PF00044; gpdh; 1.
DR Pfam; PF02800; gpdh_C; 1.
DR PRINTS; PR00078; G3PDHGRGNASE.
DR TIGRFAMS; TIGR01534; GAPDH-I; 1.
DR PROSITE; PS00071; GAPDH; 1.
DR Glycolysis; Oxidoreductase; NAD; Multigene family; Complete proteome.
FT INIT_MET 0
FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 178 178 ACTIVATES THIOL GROUP DURING CATALYSIS.
SQ SEQUENCE 334 AA; 35701 MW; 1283D3E6CF5095EC CRC64;
Query Match 87.1%; Score 27; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 VHGLK 6
Db 49 VHGLK 53

RESULT 32
G3P_BACME STANDARD; PRT; 334 AA.
AC P23722;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
GN GAP.
OS *Bacillus megaterium*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 319;
RX MEDLINE-91057128; PubMed-2123030;
RA Schlaepfer B.S., Portmann W., Branlant C., Zuber H.;
RT "Nucleotide sequence of the glyceraldehyde-3-phosphate dehydrogenase
RT from *Bacillus megaterium*.";
RL Nucleic Acids Res. 18:6422-6422(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 319;
RX MEDLINE-93083995; PubMed-1452037;
RA Schlaepfer B.S., Zuber H.;
RT "Cloning and sequencing of the genes encoding
RT glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and
RT triosephosphate isomerase (gap operon) from mesophilic *Bacillus*
RT *megaterium*: comparison with corresponding sequences from thermophilic
RT *Bacillus stearothermophilus*.";
RL Gene 122:53-62(1992).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC -!- PATHWAY: Second phase of glycolysis; first step.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54520; CAA38376.1; -;
DR EMBL; M87647; AAA73202.1; -;
DR PIR; S12696; S12696.
DR HSSP; P00362; IGD1.
DR InterPro; IPR000173; GAP_dhdrogenase.
DR InterPro; IPR006424; GAPDH-I.
DR Pfam; PF00044; gpdh; 1.
DR Pfam; PF02800; gpdh_C; 1.
DR PRINTS; PR00078; G3PDHGRGNASE.
DR TIGRFAMS; TIGR01534; GAPDH-I; 1.
DR PROSITE; PS00071; GAPDH; FALSE NEG.
DR Glycolysis; Oxidoreductase; NAD.
FT INIT_MET 0
FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 178 178 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT CONFLICT 150 150 P -> S (IN REF. 2).
SQ SEQUENCE 334 AA; 35783 MW; DEFD7C4C22532110 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 VHGLK 6
Db 49 VHGLK 53

```

RESULT 33
ADKL_ARATH
ID ADKL_ARATH STANDARD; PRT; 344 AA.
AC Q9SF85;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adenosine kinase 1 (EC 2.7.1.20) (AK 1) (Adenosine 5'-
DE phosphotransferase 1).
GS ADK1 OR AT3G09820 OR F8A24.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=cv. Columbia;
RX MEDLINE=20567825; PubMed=11115893;
RA Moffatt B.A., Wang L., Allen M.S., Stevens Y.Y., Qin W., Snider J.,
RA von Schwartzenberg K.;
RT "Adenosine kinase of Arabidopsids. Kinetic properties and gene
RT expression";
RL Plant Physiol. 124:1775-1785(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Partmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Deisny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone P., Cholsne N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehert T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Bagues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Cascuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana";
RL Nature 408:820-822(2000).
CC -!- FUNCTION: ATP DEPENDENT PHOSPHORYLATION OF ADENOSINE AND OTHER
CC RELATED NUCLEOSIDE ANALOGS TO MONOPHOSPHATE DERIVATIVES.
CC -!- CATALYTIC ACTIVITY: ATP + adenosine = ADP + AMP.
CC -!- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC -!- PATHWAY: Purine salvage. Cytokinin interconversion.
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----

```

```

DR EMBL; AF180896; AAG45248.1; -
DR EMBL; AF180894; AAG45246.1; -
DR EMBL; AC015985; AAF23253.1; -
DR HSSP; P55263; 1BX4
DR InterPro; IPR001805; Adenokinas.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00294; pfkb; 1.
DR PRINTS; PR00989; ADENOKINASE.
DR PROSITE; PS00583; PFKB_KINASES.2; 1.
DR PROSITE; PS00584; PFKB_KINASES.1; FALSE_NEG.
KW Transferase; Kinase; Purine salvage; Magnesium.
FT ACT_SITE 299 299 BY SIMILARITY.
SQ SEQUENCE 344 AA; 37836 MW; 771C789CD1D2D2E6 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVHGK 5
Db 309 LVHGK 313
|||||
RESULT 34
HN3G_HUMAN
ID HN3G_HUMAN STANDARD; PRT; 350 AA.
AC P55318; Q9UMW9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hepatocyte nuclear factor 3-gamma (HNF-3G) (Forkhead box protein A3)
DE (Fork head-related protein FKX H3).
GN FOXA3 OR HNF3G OR TCF3G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Liver;
RX MEDLINE=93271467; PubMed=8499623;
RA Hromas R., Moore J., Johnston T., Socha C., Klemsz M.;
RT "Drosophila forkhead homologues are expressed in a lineage-restricted
RT manner in human hematopoietic cells.";
RL Blood 81:2854-2859(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20359323; PubMed=10899756;
RA Navas M.A., Valsse C., Boger S., Heimesaat M., Kollee L.A.,
RA Stoffel M.;
RT "The human HNF-3 genes: cloning, partial sequence and mutation
RT screening in patients with impaired glucose homeostasis.";
RL Hum. Hered. 50:370-381(2000).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES
CC SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PEPCK, ETC.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Expressed in erythroleukemia and hepatoma cell
CC lines and in liver and pancreas. Not expressed in any other cell
CC lines or tissues examined.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; L12141; AAA58477.1; -
DR EMBL; AF176114; AAD51980.1; -
DR EMBL; AF176113; AAD51980.1; JOINED.

```

```
DR HSP; Q63245; 2HPH.
DR TRANSFAC; T02418; -.
DR Genew; HGNC:5023; FOXA3.
DR MIM; 602295; -.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Polymorphism.
FT DNA_BIND 116 207 FORK-HEAD.
FT VARIANT 91 91 G->R.
FT CONFLICT 52 52 /FTID=VAR_008859.
FT CONFLICT 83 83 MISSING (IN REF. 1).
FT CONFLICT 111 113 V -> L (IN REF. 1).
FT CONFLICT 137 137 RPL -> AP (IN REF. 1).
FT CONFLICT 156 156 M -> V (IN REF. 1).
FT CONFLICT 227 228 E -> D (IN REF. 1).
FT CONFLICT 291 291 AT -> S (IN REF. 1).
FT CONFLICT 291 291 E -> D (IN REF. 1).
SQ SEQUENCE 350 AA; 37140 MW; 07657D41B0867101 CRC64;

Query Match 87.18; Score 27; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 99 LVHGK 103

RESULT 35
MSS2_YEAST
ID MSS2_YEAST STANDARD; PRT; 351 AA.
AC P40990;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MSS2 protein.
GN MSS2 OR YDL107W OR D2340.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RX MEDLINE=95161429; PubMed=7857963;
RA Simon M., Seraphin B., Faye G.;
RT "The nuclear-encoded MSS2 gene is involved in the expression of the
RT mitochondrial cytochrome-c oxidase subunit 2 (Cox2).";
RL Biochim. Biophys. Acta 1228:95-98(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97051597; PubMed=8896274;
RA Salz J.E., Buitrago M.J., Garcia R., Revuelta J.L., del Rey F.;
RT "The sequence of a 20.3 kb DNA fragment from the left arm of
RT Saccharomyces cerevisiae chromosome IV contains the KIN28, MSS2,
RT PHO2, POL3 and DUN1 genes, and six new open reading frames.";
RL Yeast 12:1077-1084(1996).
CC -1- FUNCTION: INVOLVED IN THE EXPRESSION OF MITOCHONDRIAL CYTOCHROME C
CC OXIDASE SUBUNIT 2 (COX2).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; X74938; CAA52892.1; -.
DR PIR; C54258; C54258.

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch.)
-----
DR EMBL; X81477; CAA57230.1; -.
DR EMBL; X95644; CAA64905.1; -.
DR EMBL; Z74155; CAA98674.1; -.
DR PIR; S67649; S67649.
DR SGD; S0002265; MSS2.
DR GO; GO:0005743; C:mitochondrial inner membrane; IDA.
DR GO; GO:0017028; F:protein stabilization activity; IDA.
DR GO; GO:0006628; P:mitochondrial translocation; IDA.
DR GO; GO:0006461; P:protein complex assembly; IDA.
DR InterPro; IPR001440; TPR.
KW Mitochondrion.
KW CONFLICT 286 351 TGMEIMDLCECFGFCCVCKEENFGKARDCLCSYKVLGNDR
FT FT DKMTMINVFLESRKDSIKLLDKARL -> RVWNMT (IN
FT REF. 1).
SQ SEQUENCE 351 AA; 41476 MW; EA6E9BFC68464C67 CRC64;

Query Match 87.18; Score 27; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGLK 6
DB 194 VHGLK 198

RESULT 36
HN3G_MOUSE
ID HN3G_MOUSE STANDARD; PRT; 353 AA.
AC P35584;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 3-gamma (HNF-3G) (Forkhead box protein A3).
GN FOXA3 OR HNF3G OR TCF3G OR TCF-3G.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94307723; PubMed=8034310;
RA Kestner K., Hiemisch H., Luckow B., Schuetz G.;
RT "The HNF-3 gene family of transcription factors in mice: gene
RT structure, cDNA sequence, and mRNA distribution.";
RL Genomics 20:377-385(1994).
CC -1- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES
CC SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PEPCK, ETC.
CC INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: RESTRICTED MAINLY TO ENDODERM-DERIVED TISSUES
CC (LUNG, LIVER, STOMACH, AND SMALL INTESTINE), ALSO PRESENT
CC ADDITIONALLY IN OVARY, TESTIS, HEART, AND ADIPOSE TISSUE, BUT
CC MISSING FROM LUNG.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS AROUND DAY 15.5 OF
CC GESTATION.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; X74938; CAA52892.1; -.
DR PIR; C54258; C54258.
```



```

DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02345; -.
DR MGD; MGI:134747; Foxa3.
DR GO; GO:0001678; P:cell glucose homeostasis; IMP.
DR GO; GO:0009267; P:cellular response to starvation; IMP.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IMP.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DNA_BIND 118 209 FORK-HEAD.
SQ SEQUENCE 353 AA; 37601 MW; 28F060A8E944D5B9 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 101 LVHGK 105

RESULT 37
HN3G_RAT
ID HN3G_RAT STANDARD; PRT; 354 AA.
AC F32183;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 3-gamma (HNF-3G) (Forkhead box protein A3).
GN FOXA3 OR HNF3G OR TCF3G OR TCF-3G.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91160974; PubMed=1672118;
RA Lai E., Prezioso V.R., Tao W.F., Chen W.S., Darnell J.E. Jr.;
RT "Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head.";
RL Genes Dev. 5:416-427(1991).
RN [2]
RP STRUCTURE BY NMR OF 107-223.
RX MEDLINE=93323996; PubMed=8332212;
RA Clark K.L., Halay E.D., Lai E., Burley S.K.;
RT "Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5";
RL Nature 364:412-420(1993).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PEPCK, ETC.
CC INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L09648; AAA1339.1; -.
DR PIR; S35090; S35090.
DR HSSP; Q63245; 2HFH.

```

```

DR TRANSFAC; T01050; -.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DNA_BIND 118 209 FORK-HEAD.
SQ SEQUENCE 354 AA; 37652 MW; 318B01ECCE7C365C CRC64;

Query Match 87.1%; Score 27; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
DB 101 LVHGK 105

RESULT 38
YFBQ_HAEIN
ID YFBQ_HAEIN STANDARD; PRT; 404 AA.
AC P71348;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable aminotransferase HI0286 (EC 2.6.1.-).
GN HI0286.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES. STRONG, TO E.COLI YFBQ.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32714; AAC21948.1; -.
DR TIGR; HI0286; -.
DR InterPro; IPR001176; ACC_synthase.
DR InterPro; IPR004839; Aminotransf1/2.
DR InterPro; IPR004838; NHtransf1.
DR Pfam; PF00155; aminotran_1.2; 1.
DR PRINTS; PR00753; ACCSYNTHASE.
DR PROSITE; PS00105; AA_TRANSFER_CLASS_1; FALSE_NEG.
KW Hypothetical protein; Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.

```

FT BINDING 240 240 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 404 AA; 45160 MW; 7F2C3E24CC69FB46 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 363 LVHGK 367
|||||

RESULT 39
ARGJ_BRUME STANDARD; PRT; 413 AA.
AC Q8YJF9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Arginine biosynthesis bifunctional protein argJ [Includes: Glutamate
N-acetyltransferase (EC 2.3.1.35) (Ornithine acetyltransferase)
(Ornithine transacetylase) (OATase); Amino-acid acetyltransferase
(EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)] [Contains: Arginine
biosynthesis bifunctional protein argJ alpha chain; Arginine
biosynthesis bifunctional protein argJ beta chain].
GN ARGJ OR BMEI0124.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,
Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
Haseikorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- FUNCTION: Catalyzes two activities which are involved in the
cyclic version of arginine biosynthesis: the synthesis of
acetylglutamate from glutamate and acetyl-CoA, and of ornithine by
transacetylation between acetylornithine and glutamate (By
similarity).
CC -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + L-glutamate = L-
ornithine + N-acetyl-L-glutamate.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + L-glutamate = CoA + N-acetyl-L-
glutamate.
CC -!- PATHWAY: Arginine biosynthesis; first step.
CC -!- PATHWAY: Arginine biosynthesis; fifth step.
CC -!- SUBUNIT: Heterotetramer of two alpha and two beta chains (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- MISCELLANEOUS: Some bacteria possess a monofunctional argJ, i.e.,
capable of catalyzing only the fifth step of the arginine
biosynthetic pathway.
CC -!- SIMILARITY: Belongs to the argJ family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL; AE009455; AAL51306.1; -.
CC PIR; AG3267; AG3267.
CC HANAP; MF_01106; -; 1.
CC InterPro; IPR002813; ArgJ.

DR Pfam; PF01960; ArgJ; 1.
DR ProDom; PD004193; ArgJ; 1.
DR TIGRFAMs; TIGR00120; ArgJ; 1.
KW Arginine biosynthesis; Multifunctional enzyme; Transferase;
KW Acyltransferase; Complete proteome.
FT CHAIN 1 194
FT CHAIN 195 413
FT SITE 194 195
FT SITE 195 413
SQ SEQUENCE 413 AA; 43300 MW; BE671C6938D04DCE CRC64;

Query Match 87.1%; Score 27; DB 1; Length 413;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
Db 71 LVHGK 75
|||||

RESULT 40
ARGJ_BRUSU STANDARD; PRT; 413 AA.
AC Q8FYE2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Arginine biosynthesis bifunctional protein argJ [Includes: Glutamate
N-acetyltransferase (EC 2.3.1.35) (Ornithine acetyltransferase)
(Ornithine transacetylase) (OATase); Amino-acid acetyltransferase
(EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)] [Contains: Arginine
biosynthesis bifunctional protein argJ alpha chain; Arginine
biosynthesis bifunctional protein argJ beta chain].
GN ARGJ OR BR1941.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,
Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.W.;
RT "The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -!- FUNCTION: Catalyzes two activities which are involved in the
cyclic version of arginine biosynthesis: the synthesis of
acetylglutamate from glutamate and acetyl-CoA, and of ornithine by
transacetylation between acetylornithine and glutamate (By
similarity).
CC -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + L-glutamate = L-
ornithine + N-acetyl-L-glutamate.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + L-glutamate = CoA + N-acetyl-L-
glutamate.
CC -!- PATHWAY: Arginine biosynthesis; first step.
CC -!- PATHWAY: Arginine biosynthesis; fifth step.
CC -!- SUBUNIT: Heterotetramer of two alpha and two beta chains (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- MISCELLANEOUS: Some bacteria possess a monofunctional argJ, i.e.,
capable of catalyzing only the fifth step of the arginine
biosynthetic pathway.
CC -!- SIMILARITY: Belongs to the argJ family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL; AE009455; AAL51306.1; -.
CC PIR; AG3267; AG3267.
CC HANAP; MF_01106; -; 1.
CC InterPro; IPR002813; ArgJ.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE014483; AAN30833.1; -.
DR TIGR; BR1941; -.
DR HAMAP; MF_01106; -; 1.
DR InterPro; IPR002813; ArgJ.
DR Pfam; PF01960; ArgJ; 1.
DR ProDom; PD004193; ArgJ; 1.
DR TIGRFAMs; TIGR00120; ArgJ; 1.
DR Arginine biosynthesis; Multifunctional enzyme; Transferase;
KW Acyltransferase; Complete proteome.
FT CHAIN 1 194
FT ARGinine BIOSYNTHESIS BIFUNCTIONAL
FT PROTEIN ARGJ ALPHA CHAIN (BY SIMILARITY).
FT CHAIN 195 413
FT ARGinine BIOSYNTHESIS BIFUNCTIONAL
FT PROTEIN ARGJ BETA CHAIN (BY SIMILARITY).
FT SITE 194 195
FT CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
SQ SEQUENCE 413 AA; 43244 MW; 44C1DEF0E4A63B04 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 413;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGK 5
|
|
|
|
|
Db 71 LVHGK 75

RESULT 41
KCRU_RAT
ID KCRU_RAT STANDARD; PRT; 418 AA.
AC P25809;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Creatine kinase, ubiquitous mitochondrial precursor (EC 2.7.3.2) (U-
DE MCK) (Mia-CK) (Acidic-type mitochondrial creatine kinase).
GN CKMT1 OR CKMT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Intestine;
RX MEDLINE=91316135; PubMed=1859839;
RA Payne R.M., Haas R.C., Strauss A.W.;
RT "Structural characterization and tissue-specific expression of the
RT mRNAs encoding isoenzymes from two rat mitochondrial creatine kinase
genes";
RL Biochim. Biophys. Acta 1089:352-361(1991).
CC -1- FUNCTION: REVERSIBLY CATALYZES THE TRANSFER OF PHOSPHATE BETWEEN
CC ATP AND VARIOUS PHOSPHOGENS (E.G. CREATINE PHOSPHATE). CREATINE
CC KINASE ISOENZYMES PLAY A CENTRAL ROLE IN ENERGY TRANSDUCTION IN
CC TISSUES WITH LARGE, FLUCTUATING ENERGY DEMANDS, SUCH AS SKELETAL
CC MUSCLE, HEART, BRAIN, AND SPERMATOZOIA.
CC -1- CATALYTIC ACTIVITY: ATP + creatine -> ADP + phosphocreatine.
CC -1- SUBUNIT: EXISTS AS AN OCTAMER COMPOSED OF FOUR MTCK HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane; outer side.
CC -1- TISSUE SPECIFICITY: IN MANY TISSUES, WITH HIGHEST LEVELS IN BRAIN
CC GUT, AND KIDNEY.
CC -1- MISCELLANEOUS: MITOCHONDRIAL CREATINE KINASE BINDS CARDIOLIPIN.
CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; X59737; CAA42415.1; -.
DR PIR; S17189; S17189.
DR HSP; P11009; ICRK.
DR InterPro; IPR000749; ATP-gua_Ptrans.
DR Pfam; PF00217; ATP-gua_Ptrans; 1.
DR PROSITE; PS00112; GUANIDO_KINASE; 1.
KW Transferase; Kinase; Multigene family; Mitochondrion; Transit peptide.
FT TRANSIT 1 39
FT TRANSIT 1 39
FT CHAIN 40 418
FT CREATINE KINASE, UBIQUITOUS
FT MITOCHONDRIAL.
FT DOMAIN 40 64
FT CARDIOLIPIN-BINDING (BY SIMILARITY).
FT ACT_SITE 317 317
FT BY SIMILARITY.
SQ SEQUENCE 418 AA; 47028 MW; 468339C52E4232D5 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 418;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGK 5
|
|
|
|
|
Db 413 LVHGK 417

RESULT 42
LE21_ARCFU
ID LE21_ARCFU STANDARD; PRT; 418 AA.
AC O28316;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase large subunit 1 (EC 4.2.1.33)
DE (Isopropylmalate isomerase 1) (Alpha-IPM isomerase 1) (IPMI 1).
GN LEUC1 OR AFI963.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.W., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus";
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-
CC isopropylmaleate.
CC -1- COFACTOR: Binds 1 4Fe-4S cluster per subunit (By similarity).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Heterodimer of leuc and leud (By similarity).
CC -1- SIMILARITY: Belongs to the aconitase/IPM isomerase family. Leuc 2
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF000967; AAB89290.1; ALT_INIT.
 DR TIGR: AF1963; .
 DR HAMAP: ME_01027; .
 DR InterPro: IPR001030; Aconitase_N.
 DR pfam: PF006251; Cis-H_aconitase.
 DR PRINTS: PR00330; aconitase.1.
 DR ProDom: PD000511; Aconitase_N.
 DR TIGRFAMS: TIGR01343; haca_fam; 1.
 DR PROSITE: PS00450; ACONITASE.1; 1.
 DR PROSITE: PS01244; ACONITASE.2; 1.
 KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.
 FT METAL 298 298 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 358 358 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 361 361 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 418 AA; 45140 MW; 51B2BD0A3D9C9A22 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6
 Db 163 VHGL 167
 |||||

RESULT 43
 ID SYI_BUCAI
 AC P57221; STANDARD; PRT; 422 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tyrosyl-tRNA synthetase (EC 6.1.1.1) (Tyrosine--tRNA ligase) (TyrRS).
 TYRS OR BU121.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -!- CATALYTIC ACTIVITY: ATP + L-tyrosine + tRNA(Tyr) = AMP +
 CC Diphosphate + L-tyrosyl-tRNA(Tyr).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminocyl-tRNA synthetase family.
 CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AP001118; BAB12839.1; .
 DR HSSP: P00952; lTYC.
 DR InterPro: IPR002942; S4.
 DR InterPro: IPR002305; tRNA-synt_1b.
 DR InterPro: IPR001412; tRNA-synt_1.

DR InterPro: IPR002307; Tyr_tRNA-synt_1b.
 DR pfam: PF00579; tRNA-synt_1b; 1.
 DR PRINTS: PR01040; TRNASYNTHYR.
 DR SMART: SM00363; S4; 1.
 DR TIGRFAMS: TIGR00234; tyrs; 1.
 DR PROSITE: PS00178; AA_tRNA_LIGASE_I; FALSE_NEG.
 DR PROSITE: PS50889; S4; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 42 51 "HIGH" REGION.
 FT SITE 235 239 "KMSKS" REGION.
 FT BINDING 238 238 ATP (BY SIMILARITY).
 FT DOMAIN 357 414 S4 RNA-BINDING.
 SQ SEQUENCE 422 AA; 48675 MW; D83C40982D041048 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 422;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGK 5
 Db 310 LVHGK 314
 |||||

RESULT 44
 ID YA54_HAEIN
 AC P44104; STANDARD; PRT; 460 AA.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein HI1054.
 GN HI1054.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerkavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd";
 RL Science 269:496-512(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U32786; AAC22713.1; .
 DR PIR: E64019; E64019.
 DR TIGR: HI1054; .
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 460 AA; 53471 MW; 0D80EA220AFDEDA8 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 460;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VHGL 6

```

Db          147 VHGL 151
|||||
RESULT 45
SYK_HELPJ
ID  SYK_HELPJ      STANDARD;      PRT;      501 AA.
AC  Q9ZMP8;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
GN  LYSS OR JHP0170.
OS  Helicobacter pylori J99 (Campylobacter pylori J99).
OC  Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OX  Helicobacteraceae; Helicobacter.
NCBI_TaxID=85963;
[.]
SEQUENCE FROM N.A.
RP  MEDLINE=99120557; PubMed=9923682;
RX  Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA  Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA  Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA  Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA  Trust T.J.;
RT  "Genomic sequence comparison of two unrelated isolates of the human
RL  gastric pathogen Helicobacter pylori.";
RL  Nature 397:176-180(1999).
CC  -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC  + L-lysyl-tRNA(Lys).
CC  -!- COFACTOR: Binds 3 magnesium ions per subunit (By similarity).
CC  -!- SUBUNIT: Homodimer (By similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; AE001455; AAD05751.1; -.
DR  PIR; F71965; F71965.
DR  HSSP; P14825; 1E10.
DR  HAMAP; MF_00252; -.
DR  InterPro; IPR004364; tRNA-synt_2.
DR  InterPro; IPR002313; tRNA-synt_lys_2.
DR  InterPro; IPR004365; tRNA_anti.
DR  InterPro; IPR006195; tRNA_ligase_II.
DR  Pfam; PF00152; tRNA-synt_2; 1.
DR  Pfam; PF01336; tRNA_anti; 1.
DR  PRINTS; PR00982; TRNASYNTHLYS.
DR  TIGRFAMs; TIGR00499; lyss_bact; 1.
DR  PROSITE; PS00862; AA_TRNA_LIGASE_II; 1.
KW  Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW  Metal-binding; Magnesium; Complete proteome.
FT  METAL 402 402 MAGNESIUM 1 (BY SIMILARITY).
FT  METAL 409 409 MAGNESIUM 1 AND 2 (BY SIMILARITY).
SQ  SEQUENCE 501 AA; 57667 MW; 312B4381708A00F4 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 501;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LVHGKL 6
| ||||
Db 353 LTHGKL 358

```

Search completed: October 14, 2003, 13:12:22
Job time : 10.6364 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:21 ; Search time 32.7273 Seconds
(without alignments)
47.310 Million cell updates/sec

Title: US-09-856-050-19_COPY_24_29
Perfect score: 31
Sequence: 1 LVHGKL 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	31	100.0	199	10	Q9SFU5	Q9sfu5 arabidopsis
2	31	100.0	384	16	Q8PND5	Q8pnd5 xanthomonas
3	31	100.0	384	16	Q8PBT7	Q8pbt7 xanthomonas
4	31	100.0	385	2	Q937N9	Q937n9 alcaligenes
5	31	100.0	387	16	Q8XTI7	Q8xti7 ralstonia s
6	31	100.0	388	2	Q8VPS9	Q8vps9 burkholderi
7	31	100.0	489	10	Q9LWL1	Q9lwl1 oryza sativ
8	31	100.0	553	16	Q8Z0A6	Q8z0a6 anabaena sp
9	31	100.0	633	16	Q9F3Q3	Q9f3q3 streptomyce
10	31	100.0	767	10	Q9SVR4	Q9svr4 arabidopsis
11	31	100.0	1274	10	Q9LVC5	Q9lvc5 arabidopsis
12	30	96.8	111	5	Q8BMK7	Q8bmk7 brugia mala
13	30	96.8	220	5	O44932	O44932 brugia mala
14	30	96.8	389	16	Q8Z904	Q8z904 salmonella
15	30	96.8	389	16	Q8X694	Q8x694 escherichia
16	30	96.8	389	16	Q8FKH2	Q8fkh2 escherichia

17	29	93.5	107	12	Q8QO14	Q8qq14 camelpox vi
18	29	93.5	127	12	Q98409	Q98409 newcastle d
19	29	93.5	128	10	Q95WH6	Q95wh6 zea mays (m
20	29	93.5	143	10	O24582	O24582 zea mays (m
21	29	93.5	265	16	Q9K7Q5	Q9k7q5 bacillus ha
22	29	93.5	372	16	Q8D954	Q8d954 vibrio vuln
23	29	93.5	393	16	Q9KRT4	Q9krt4 vibrio chol
24	29	93.5	467	5	Q8SR98	Q8sr98 encephalito
25	29	93.5	518	2	P95535	P95535 alcaligenes
26	29	93.5	518	2	Q8GPH4	Q8gph4 achromobact
27	29	93.5	676	2	Q9EZA2	Q9eza2 zymomonas m
28	29	93.5	685	4	O81WK5	O81wk5 homo sapien
29	29	93.5	724	16	Q8EZB0	Q8ezb0 leptospira
30	29	93.5	926	4	Q8TE49	Q8tte49 homo sapien
31	29	93.5	926	11	Q8R554	Q8r554 mus musculu
32	29	93.5	980	5	Q17592	Q17592 caenorhabdi
33	29	93.5	1446	12	Q8BDE1	Q8bde1 acute bee p
34	29	93.5	1446	12	Q8BDE0	Q8bde0 acute bee p
35	28	90.3	38	4	Q16481	Q16481 homo sapien
36	28	90.3	78	2	Q8RL65	Q8rl65 pseudomonas
37	28	90.3	136	16	Q8DHG6	Q8dhg6 synecococc
38	28	90.3	178	10	Q9ZUD3	Q9zud3 arabidopsis
39	28	90.3	188	16	Q97F26	Q97f26 clostridium
40	28	90.3	241	11	Q8R0C2	Q8r0c2 mus musculu
41	28	90.3	252	10	Q8GSK7	Q8gsk7 oryza sativ
42	28	90.3	254	16	Q9K7R5	Q9k7r5 bacillus ha
43	28	90.3	278	5	Q9V4S8	Q9vas8 drosophila
44	28	90.3	287	10	O04627	O04627 arabidopsis
45	28	90.3	312	16	Q8F7L2	Q8f7l2 leptospira

ALIGNMENTS

RESULT 1

Q9SFU5 PRELIMINARY; PRT; 199 AA.
ID Q9SFU5;
AC Q9SFU5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T1B9.19 protein.
GN T1B9.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T1B9 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC012395; AAF20231.1; .
SQ SEQUENCE 199 AA; 21843 MW; F4E844AE5B2D5D57 CRC64;

Query Match 100.0%; Score 31; DB 10; Length 199;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
|||||
Db 90 LVHGKL 95

RESULT 2

Q8PND5 PRELIMINARY; PRT; 384 AA.
ID Q8PND5;
AC Q8PND5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Citrate synthase 2.
GN	PRPC OR XAC1138.
OS	Xanthomonas axonopodis (pv. citri).
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC	Xanthomonadaceae; Xanthomonas.
OX	NCBI_TaxID=92829;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=306 / ATCC 13902 / XV 101;
RX	MEDLINE=22022145; PubMed=12024217;
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA	Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA	Cicarella R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA	Pereira H.A., Rossi A.A., Sena J.A.D., Silva C., de Souza R.F.,
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA	Sutubal J.C., Kitajima J.P.;
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing
RT	host specificities."
RL	Nature 417:459-463(2002).
DR	EMBL; AE011742; AA36010.1; -.
DR	InterPro; IPR002020; Citrate_synt.
DR	Pfam; PF00285; citrate_synt; 1.
DR	PRINTS; PR00143; CITRGSYNTHASE.
DR	PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR	Complete proteome.
SW	SEQUENCE 384 AA; 42381 MW; 79F7EF5E76EBDCAC CRC64;
QY	1 LVHGKL 6
DB	
DB	60 LVHGKL 65
RESULT 3	
Q8PB7	ID Q8PB7 PRELIMINARY; PRT; 384 AA.
AC	Q8PB7;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Citrate synthase 2.
GN	PRPC OR XCC1032.
OS	Xanthomonas campestris (pv. campestris).
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC	Xanthomonadaceae; Xanthomonas.
OX	NCBI_TaxID=340;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 33913 / NCPPB 528;
RX	MEDLINE=22022145; PubMed=12024217;
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA	Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA	Cicarella R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA	Pereira H.A., Rossi A.A., Sena J.A.D., Silva C., de Souza R.F.,
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA	Sutubal J.C., Kitajima J.P.;
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing
RT	host specificities."
RL	Nature 417:459-463(2002).
DR	EMBL; AE011742; AA36010.1; -.
DR	InterPro; IPR002020; Citrate_synt.
DR	Pfam; PF00285; citrate_synt; 1.
DR	PRINTS; PR00143; CITRGSYNTHASE.
DR	PROSITE; PS00480; CITRATE_SYNTHASE; 1.
DR	Complete proteome.
SW	SEQUENCE 384 AA; 42381 MW; 79F7EF5E76EBDCAC CRC64;
QY	1 LVHGKL 6
DB	
DB	60 LVHGKL 65
Query Match	100.0%; Score 31; DB 16; Length 384;
Best Local Similarity	100.0%; Pred. No. 1.le+02;
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

Q8XTI7
ID Q8XTI7 PRELIMINARY; PRT; 387 AA.
AC Q8XTI7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable methylcitrate synthase (Citrate synthase 2) protein
DE (EC 4.1.3.7).
GN PRPC OR RSP0121 OR RS03001.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646076; CAD17272.1; -.
DR InterPro: IPR002020; Citrate_synt.
DR Pfam: PF00285; citrate_synt; 1.
DR PRINTS; PRO0143; CITRISYNTHASE.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
KW Lyase; Plasmid; Complete proteome.
SQ SEQUENCE 387 AA; 42634 MW; 2C7365DFA5CE81F1 CRC64;

Query Match 100.0%; Score 31; DB 16; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 64 LVHGKL 69
|||||

RESULT 6
Q8VPS9
ID Q8VPS9 PRELIMINARY; PRT; 388 AA.
AC Q8VPS9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 2-methylcitrate synthase (EC 4.1.3.7).
GN PRPC.
OS Burkholderia sacchari.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=159450;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IPT101T;
RX MEDLINE=21633825; PubMed=11772636;
RA Bramer C.O., Silva L.F., Gomez J.G.C., Priefert H., Steinbuechel A.;
RT "Identification of the 2-Methylcitrate Pathway Involved in the
RT Catabolism of Propionate in the Polyhydroxyalkanoate-Producing Strain
RT Burkholderia sacchari IPT101(T) and Analysis of a Mutant Accumulating
RT a Copolyester with Higher 3-Hydroxyvalerate Content.";
RL Appl. Environ. Microbiol. 68:271-279(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IPT101T;
RA Bramer C.O., Steinbuechel A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
OX OXALOACETATE.

```

```

CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
DR EMBL; AY033092; AAK52340.1; -.
DR InterPro: IPR002020; Citrate_synt.
DR Pfam: PF00285; citrate_synt; 1.
DR PRINTS; PRO0143; CITRISYNTHASE.
DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle.
SQ SEQUENCE 388 AA; 42690 MW; 2C0B52CAAA2549FC CRC64;

Query Match 100.0%; Score 31; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 65 LVHGKL 70
|||||

RESULT 7
Q9LWL1
ID Q9LWL1 PRELIMINARY; PRT; 489 AA.
AC Q9LWL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0485D09.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001859; BAA94778.1; -.
DR Gramene; Q9LWL1; -.
DR InterPro: IPR001084; Crystallin.
DR Pfam: PF00400; WD40; 3.
DR SMART; SM00320; WD40; 3.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 489 AA; 52544 MW; 5A15CEF790535342 CRC64;

Query Match 100.0%; Score 31; DB 10; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
DB 335 LVHGKL 340
|||||

RESULT 8
Q8Z0A6
ID Q8Z0A6 PRELIMINARY; PRT; 553 AA.
AC Q8Z0A6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Serine/threonine Kinase.
GN AL01192.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003581; BAB77716.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS0108; PROTEIN_KINASE_ST; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 553 AA; 62372 MW; 6959D0C98C866623 CRC64;

Query Match 100.0%; Score 31; DB 16; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 155 LVHGKL 160
|||||

RESULT 9
Q9F303
ID Q9F303 PRELIMINARY; PRT; 633 AA.
AC Q9F303;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SC07634.
GN SC07634 OR SC10F4.07.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL939132; CAC16967.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 633 AA; 69716 MW; 9F2F3012548701E3 CRC64;

Query Match 100.0%; Score 31; DB 16; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 602 LVHGKL 607
|||||

RESULT 10
Q9SVR4
ID Q9SVR4 PRELIMINARY; PRT; 767 AA.

```

```

AC Q9SVR4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 89.6 kDa protein.
GN F17N18.20 OR AT4G13130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Mannhaupt G., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049751; CAB41923.1; -.
DR EMBL; AL161535; CAB78355.1; -.
DR InterPro; IPR004146; DCL.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF03107; DCL; 3.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 767 AA; 89625 MW; 37A2D510FE1A8A9F CRC64;

Query Match 100.0%; Score 31; DB 10; Length 767;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 196 LVHGKL 201
|||||

RESULT 11
Q9LVQ5
ID Q9LVQ5 PRELIMINARY; PRT; 1274 AA.
AC Q9LVQ5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similarity to salt-inducible protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
DR EMBL; AB018120; BRA97283.1; -.
DR InterPro; IPR002110; ANK.

```

DR InterPro; IPR002885; PPR.
 DR Pfam; PF01535; PPR; 26.
 DR SMART; SM00248; ANK; 1.
 DR TIGRFAMs; TIGR00756; PPR; 26.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 1274 AA; 142332 MW; 3AA044E3C35405B9 CRC64;
 Query Match 100.0%; Score 31; DB 10; Length 1274;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 47 LVHGKL 52
 RESULT 12
 Q9BMK7 PRELIMINARY; PRT; 111 AA.
 AC Q9BMK7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Vespid allergen antigen-like protein (Fragment).
 OS Brugia malayi.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Brugia.
 OX NCBI_TaxID=6279;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rao K.V.N., Munirathnam G., Kalyanasundaram R.;
 RT "Putative protective antigens from filarial parasite Brugia malayi
 RT selected using phage display technique.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF319996; AAK07633.1;
 DR InterPro; IPR001283; Allrgn_V5/Tpx1.
 DR Pfam; PF00188; SCP; 1.
 DR ProDom; PD000542; Allrgn_V5/Tpx1; 1.
 DR SMART; SM00198; SCP; 1.
 FT NON_TER 111 111
 SQ SEQUENCE 111 AA; 12804 MW; FE174C09D0D081FF CRC64;
 Query Match 96.8%; Score 30; DB 5; Length 111;
 Best Local Similarity 83.3%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 43 LVHGKL 48
 RESULT 13
 O44932 PRELIMINARY; PRT; 220 AA.
 ID O44932;
 AC O44932;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Vespid allergen antigen homolog (Venom allergen antigen-like protein
 DE 1).
 OS Brugia malayi.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Brugia.
 OX NCBI_TaxID=6279;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McCarthy J.S., Hopkins R.M.;
 RT "Cloning of the Brugia malayi VA homolog.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Murray J., Gregory W.F., Atmadja A.K., Maizels R.M.;

RT "Expression and immune recognition of Brugia malayi VAL-1, a homolog
 RT of vespid venom allergens and Ankylostoma secreted proteins.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF042088; AAB97283.2;
 DR EMBL; AF334661; AAK12274.1;
 DR InterPro; IPR001283; Allrgn_V5/Tpx1.
 DR Pfam; PF00188; SCP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR ProDom; PD000542; Allrgn_V5/Tpx1; 1.
 DR SMART; SM00198; SCP; 1.
 SQ SEQUENCE 220 AA; 24761 MW; F22F11AD7768FE65 CRC64;
 Query Match 96.8%; Score 30; DB 5; Length 220;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 43 LVHGKL 48
 RESULT 14
 Q8Z904 PRELIMINARY; PRT; 389 AA.
 ID Q8Z904;
 AC Q8Z904;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Methylcitrate synthase (EC 4.1.3.-).
 GN PRPC OR STV0401.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL; AL627266; CAD08824.1;
 DR InterPro; IPR002020; Citrate_synt.
 DR Pfam; PF00285; Citrate_synt; 1.
 DR PRINTS; PR00143; CITRATESYNTHASE.
 DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
 KW Lyase; Complete proteome.
 SQ SEQUENCE 389 AA; 43212 MW; 0EB9091EAE5DE8D4 CRC64;
 Query Match 96.8%; Score 30; DB 16; Length 389;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 65 LVHGKL 70
 RESULT 15
 Q8X694 PRELIMINARY; PRT; 389 AA.
 ID Q8X694;
 AC Q8X694;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

Putative citrate synthase, propionate metabolism?
 PRPC OR Z0428 OR ECS0386.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamianos K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AE005212; AAG54681.1; -;
 DR EMBL; AP002551; BAB33809.1; -;
 DR InterPro; IPR002020; Citrate_synt.
 DR Pfam; PF00285; citrate_synt; 1.
 DR PRINTS; PR00143; CITRTSNTHASE.
 DR PROSITE; PS00480; CITRATE_SYNTHASE; 1.
 KW Complete proteome.
 SQ SEQUENCE 389 AA; 43107 MW; 953E4093CD11F624 CRC64;
 Query Match 96.8%; Score 30; DB 16; Length 389;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 65 LIHGKL 70
 RESULT 16
 Q8FKH2 PRELIMINARY; PRT; 389 AA.
 ID Q8FKH2
 AC Q8FKH2
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 2-methylcitrate synthase (EC 4.1.3.31).
 GN PRPC OR C0452.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016756; AAN78931.1; -;
 KW Lyase; Complete proteome.

SQ SEQUENCE 389 AA; 43076 MW; 949CC2ACC1E1E8EB CRC64;
 Query Match 96.8%; Score 30; DB 16; Length 389;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 65 LIHGKL 70
 RESULT 17
 Q8QQ14 PRELIMINARY; PRT; 107 AA.
 ID Q8QQ14
 AC Q8QQ14
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE CPM175.5L.
 GN CPM175.5L.
 OS Camelpox virus (strain CP-1).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=203174;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMS;
 RX PubMed=11907336;
 RA Gubser C., Smith G.L.;
 RT "The sequence of camelpox virus shows it is most closely related to
 variola virus, the cause of smallpox";
 RL J. Gen. Virol. 83:855-872(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMS;
 RA Gubser C., Smith G.L.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY009089; AAG37678.1; -;
 SQ SEQUENCE 107 AA; 11803 MW; F5315BF6096FEF4F CRC64;
 Query Match 93.5%; Score 29; DB 12; Length 107;
 Best Local Similarity 83.3%; Pred. No. 79;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LVHGKL 6
 Db 43 LVHGKI 48
 RESULT 18
 Q98409 PRELIMINARY; PRT; 127 AA.
 ID Q98409
 AC Q98409
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hemagglutinin-neuraminidase (Fragment).
 OS Newcastle disease virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW/95-9;
 RX MEDLINE=97345034; PubMed=9201401;
 RA Yang C.-Y., Chang P.-C., Hwang J.-M., Shieh H.K.;
 RT "Nucleotide sequence and phylogenetic analysis of Newcastle disease
 virus isolates from recent outbreaks in Taiwan";
 RL Avian Dis. 41:365-373(1997).
 DR EMBL; U75714; AAB18357.1; -;
 DR InterPro; IPR000665; Hem-neuramndse.
 DR Pfam; PF00423; HN; 1.
 FT NON_TER 1

```
FT NON_TER 127 127
SQ SEQUENCE 127 AA; 14353 MW; 72E2AE35572F4C4 CRC64;

Query Match 93.5%; Score 29; DB 12; Length 127;
Best Local Similarity 83.3%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 26 MVHGKL 31

RESULT 19
Q9SWH6 PRELIMINARY; PRT; 128 AA.
AC Q9SWH6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 14.1 kDa protein (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Stapleton A.E., Walbot V.;
RT "Maize clones with putative DNA-binding activity.";
CC Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
DR EMBL; AF152601; AAD45283.1; -.
DR HSSP; P04410; 1A25.
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 1.
DR SMART; SM00239; C2; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
KW Hypothetical protein.
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14053 MW; 84A758CF64658B23 CRC64;

Query Match 93.5%; Score 29; DB 10; Length 128;
Best Local Similarity 83.3%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 1 MVHGKL 6

RESULT 20
O24582 PRELIMINARY; PRT; 143 AA.
AC O24582;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Novel protein...
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B73;
RA Betawar N.M., Baysdorfer C.;
RT "Novel maize gene.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
DR EMBL; U04437; AAB06331.1; -.
DR HSSP; P04410; 1A25.
```

```
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 1.
DR SMART; SM00239; C2; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
SQ SEQUENCE 143 AA; 15856 MW; 0353FC5B225F7F43 CRC64;

Query Match 93.5%; Score 29; DB 10; Length 143;
Best Local Similarity 83.3%; Pred. No. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 1 MVHGKL 6

RESULT 21
Q9K7Q5 PRELIMINARY; PRT; 265 AA.
AC Q9K7Q5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein BH3306.
GN BH3306.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP01518; BAB07025.1; -.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 265 AA; 29799 MW; 2FCB8470A876346C CRC64;

Query Match 93.5%; Score 29; DB 16; Length 265;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 198 IVHGKL 203

RESULT 22
Q8D954 PRELIMINARY; PRT; 372 AA.
ID Q8D954;
AC Q8D954;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ABC-type sugar transport system, ATPase component.
GN VV13752.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
```

```

DR EMBL; AE016806; AA011096.1; -.
KW Sugar transport; Complete proteome.
SQ SEQUENCE 372 AA; 41140 MW; 8C6737B460146227 CRC64;

Query Match 93.5%; Score 29; DB 16; Length 372;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 305 LVHGKI 310

RESULT 23
Q9KRT4
ID Q9KRT4 PRELIMINARY; PRT; 393 AA.
AC Q9KRT4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glycerol-3-phosphate ABC transporter, ATP-binding protein.
GN VC1552.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004233; AAF94706.1; -.
DR TIGR; VC1552; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 393 AA; 43508 MW; 1343D88DF0917E99 CRC64;

Query Match 93.5%; Score 29; DB 16; Length 393;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 326 LVHGKI 331

RESULT 24
Q8SR98
ID Q8SR98 PRELIMINARY; PRT; 467 AA.
AC Q8SR98;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SNF1-related protein kinase.
GN ECU08_1480.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL590448; CAD26452.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 467 AA; 53811 MW; 91156FFB272A0320 CRC64;

Query Match 93.5%; Score 29; DB 5; Length 467;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 195 MVHGKL 200

RESULT 25
P95535
ID P95535 PRELIMINARY; PRT; 518 AA.
AC P95535; P95536;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transposase.
GN ISTA.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96084387; PubMed=8526487;
RA Ogawa N., Miyashita K.;
RT "Recombination of a 3-chlorobenzoate catabolic plasmid from
RT Alkaligenes eutrophus NH9 mediated by direct repeat elements.";
RL Appl. Environ. Microbiol. 61:3788-3795(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99124613; PubMed=9925607;
RA Ogawa N., Miyashita K.;
RT "The chlorocatechol-catabolic transposon Tn5707 of Alkaligenes
RT eutrophus NH9, carrying a gene cluster highly homologous to that in
RT the 1,2,4-trichlorobenzene-degrading bacterium Pseudomonas sp. strain
RT P51, confers the ability to grow on 3-chlorobenzoate.";
RL Appl. Environ. Microbiol. 65:724-731(1999).
DR EMBL; D64144; BAA33969.1; -.
DR InterPro; IPR007101; IS21_HTH.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
DR PROSITE; PSS0531; IS21_TRANSPOSASE; 1.
KW Plasmid.

```

SQ SEQUENCE 518 AA; 58879 MW; 661A750FCD6C763A CRC64;

Query Match 93.5%; Score 29; DB 2; Length 518;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 285 IVHGKL 290

RESULT 26

Q8GFH4 ID Q8GFH4 PRELIMINARY; PRT; 518 AA.
AC Q8GFH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ista protein.
GN Ista.
OS Achromobacter xylosoxidans.
OG Plasmid pA81.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Achromobacter.
OX NCBI_TaxID=85698;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=A8;
RA Jencova V., Strnad H., Chodora Z., Ulbrich P., Hickey W.J., Paces V.;
RT "Chlorocatechol catabolic enzymes from Achromobacter xylosoxidans
RT A8";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ515144; CAD56213.1; -
KW Plasmid.
SQ SEQUENCE 518 AA; 58879 MW; 661A750FCD6C763A CRC64;

Query Match 93.5%; Score 29; DB 2; Length 518;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 285 IVHGKL 290

RESULT 27

Q9EZA2 ID Q9EZA2 PRELIMINARY; PRT; 676 AA.
AC Q9EZA2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ATP-dependent DNA helicase RecG.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Zymomonas.
OX NCBI_TaxID=542;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Shin I.S., Kang H.S.;
RT "Zymomonas mobilis ZM4 fosmid clone 43E12 complete sequence.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300471; AAG42404.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICCC; 1.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 676 AA; 75018 MW; 5C97AC5B87F98ECD CRC64;

Query Match 93.5%; Score 29; DB 2; Length 676;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 514 LVHGKM 519

RESULT 28

Q81WK5 ID Q81WK5 PRELIMINARY; PRT; 685 AA.
AC Q81WK5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to cezanne 2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035668; AAH35668.1; -
FT NON_TER 685
SQ SEQUENCE 685 AA; 76091 MW; B60C5CFE6D1995CD CRC64;

Query Match 93.5%; Score 29; DB 4; Length 685;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 545 LVHGKM 550

RESULT 29

Q8EZB0 ID Q8EZB0 PRELIMINARY; PRT; 724 AA.
AC Q8EZB0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent DNA helicase recG (EC 3.6.1.-).
GN REC OR LA3945.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011550; AAN51143.1; -
KW Helicase; Hydrolase; Complete proteome.
SQ SEQUENCE 724 AA; 82558 MW; D6AB77D67554DF6A CRC64;

Query Match 93.5%; Score 29; DB 16; Length 724;
Best Local Similarity 83.3%; Pred. No. 5.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
:|||||
Db 554 LVHGKM 559

RESULT 30

Q8TE49 ID Q8TE49 PRELIMINARY; PRT; 926 AA.
AC Q8TE49;

```
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cezanne 2 protein.
GN CEZANNE 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Evans P.C., Coadwell W.J., Kilshaw P.J.;
RT "Isolation of a novel human gene, Cezanne 2.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430383; CAD23047.1; -.
DR InterPro; IPR003323; OTU.
DR PROSITE; PS0802; OTU; 1.
SQ SEQUENCE 926 AA; 100676 MW; 6E4623CEB2C8058 CRC64;

Query Match          93.5%; Score 29; DB 4; Length 926;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 538 LVHGKM 543

RESULT 31
Q8R554
ID Q8R554 PRELIMINARY; PRT; 926 AA.
AC Q8R554;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Cezanne 2 protein.
GN AJ430384 OR CEZANNE 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Evans P.C., Coadwell W.J., Kilshaw P.J.;
RT "Isolation of a novel murine gene, Cezanne 2.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430384; CAD23048.1; -.
DR MGD; MGI:2158505; AJ430384.
DR InterPro; IPR003323; OTU.
DR PROSITE; PS0802; OTU; 1.
SQ SEQUENCE 926 AA; 100796 MW; 4D6BD05A0410BED9 CRC64;

Query Match          93.5%; Score 29; DB 11; Length 926;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 541 LVHGKM 546

RESULT 32
Q17592
ID Q17592 PRELIMINARY; PRT; 980 AA.
AC Q17592;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical 112.8 kDa protein.
GN C0266.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

```
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R., Bentley D., Kemp K., Scheet P.;
RT "The sequence of C. elegans cosmid C02G6.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55372; AAA98001.1; -.
DR MEROPS; M16.002; -.
DR WormPep; C02G6.1; CE06755.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 980 AA; 112806 MW; 356ADE82F25EDE29 CRC64;

Query Match          93.5%; Score 29; DB 5; Length 980;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 913 LVHGKI 918

RESULT 33
Q8BDE1
ID Q8BDE1 PRELIMINARY; PRT; 1446 AA.
AC Q8BDE1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Non-structural protein (Fragment).
OS Acute bee paralysis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Cricket paralysis-like viruses.
OX NCBI_TaxID=92444;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hungary 1;
RA Bakonyi T., Szendroi A., Topolska G., Rusvai M.;
RT "Nucleotide sequence analysis of the non-structural protein gene
region of ABPV strains.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF486072; AAN63803.1; -.
FT NON_TER 1
FT NON_TER 1446
SQ SEQUENCE 1446 AA; 165546 MW; 46A55449854A4C5B CRC64;

Query Match          93.5%; Score 29; DB 12; Length 1446;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 1261 LVHGKI 1266
```



```
RESULT 34
Q8BDE0 ID Q8BDE0 PRELIMINARY; PRT; 1446 AA.
AC Q8BDE0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Non-structural protein (Fragment).
OS Acute bee paralysis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Cricket paralysis-like viruses.
OX NCBI_TaxID=92444;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Poland 1;
RA Bakonyi T., Szendroi A., Topolska G., Rusvai M.;
RT "Nucleotide sequence analysis of the non-structural protein gene
RT region of ABPV strains.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF486073; AAN63804.1; -.
FT NON_TER 1
FT NON_TER 1446 1446
SQ SEQUENCE 1446 AA; 165585 MW; C5CE9BE99B3C13F CRC64;

Query Match 93.5%; Score 29; DB 12; Length 1446;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 1261 LVHGKI 1266

RESULT 35
Q16481 ID Q16481 PRELIMINARY; PRT; 38 AA.
AC Q16481;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE Aromatase protein (Fragment).
GN AROMATASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92359115; PubMed=1496995;
RA Harada N., Ogawa H., Shozu M., Yamada K.;
RT "Genetic studies to characterize the origin of the mutation in
RT placental aromatase deficiency.";
RL Am. J. Hum. Genet. 51:666-672(1992).
DR EMBL; S41951; AADI4894.2; -.
FT NON_TER 1
FT NON_TER 38 38
SQ SEQUENCE 38 AA; 4313 MW; 3FC608ADB8B8970A CRC64;

Query Match 90.3%; Score 28; DB 4; Length 38;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 12 LMHGKL 17

RESULT 36
Q8RL65 ID Q8RL65 PRELIMINARY; PRT; 78 AA.
AC Q8RL65;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
```

```
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Macp14.
GN WACP14.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=NCIMB 10586;
RA El-Sayed A.K., Hotherhall J., Cooper S.M., Stephens E., Thomas C.M.;
RT "Nucleotide sequence of a 75 kb region of the chromosome of
RT Pseudomonas fluorescens NCIMB 10586 required for biosynthesis of the
RT polyketide antibiotic mupirocin.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF318063; AAM12920.1; -.
DR InterPro; IPR006163; Pp_bind.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
KW Phosphopantetheine.
SQ SEQUENCE 78 AA; 8608 MW; 1C5DE3D892CF5B76 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 78;
Best Local Similarity 83.3%; Pred. No. 96;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 73 LVHGKL 78

RESULT 37
Q8DHG6 ID Q8DHG6 PRELIMINARY; PRT; 136 AA.
AC Q8DHG6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tlr1993 protein.
GN TLR1993.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005375; BAC09545.1; -.
KW Complete proteome.
SQ SEQUENCE 136 AA; 16015 MW; 5931250EBA4A37E6 CRC64;

Query Match 90.3%; Score 28; DB 16; Length 136;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 108 MHGKL 113

RESULT 38
Q9ZUD3 ID Q9ZUD3 PRELIMINARY; PRT; 178 AA.
AC Q9ZUD3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
```

```

01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
F508.17 protein.
GN F508.17
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,
RA Li J., Kremenetskaia I., Luros J., Altafi H., Gonzalez A., Araujo R.,
RA Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Huizar L.,
RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F508 sequence.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005950; AAC98017.1;
DR InterPro; IPR003863; DUF220.
DR Pfam; PF02713; DUF220; 1.
SQ SEQUENCE 178 AA; 20314 MW; 2FC3556C05DFD1FA CRC64;

Query Match 90.3%; Score 28; DB 10; Length 178;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 7 LLHGKL 12

RESULT 39
Q97F26
ID Q97F26 PRELIMINARY; PRT; 188 AA.
AC Q97F26;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Predicted membrane protein.
GN CAC2929.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Gossling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007790; AAK80871.1;
KW Complete proteome.
SQ SEQUENCE 188 AA; 21127 MW; 8B7875FF42CF3D3E CRC64;

Query Match 90.3%; Score 28; DB 16; Length 188;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 136 VHGKL 141

01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
F508.17 protein.
GN F508.17
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,
RA Li J., Kremenetskaia I., Luros J., Altafi H., Gonzalez A., Araujo R.,
RA Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Huizar L.,
RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F508 sequence.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005950; AAC98017.1;
DR InterPro; IPR003863; DUF220.
DR Pfam; PF02713; DUF220; 1.
SQ SEQUENCE 178 AA; 20314 MW; 2FC3556C05DFD1FA CRC64;

Query Match 90.3%; Score 28; DB 10; Length 178;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 7 LLHGKL 12

RESULT 39
Q97F26
ID Q97F26 PRELIMINARY; PRT; 188 AA.
AC Q97F26;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Predicted membrane protein.
GN CAC2929.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Gossling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007790; AAK80871.1;
KW Complete proteome.
SQ SEQUENCE 188 AA; 21127 MW; 8B7875FF42CF3D3E CRC64;

Query Match 90.3%; Score 28; DB 16; Length 188;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 136 VHGKL 141

```

```

RESULT 40
Q8ROC2
ID Q8ROC2 PRELIMINARY; PRT; 241 AA.
AC Q8ROC2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 27.0 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027078; AAH27078.1;
DR InterPro; IPR000175; Na/ntran_sympot.
DR Pfam; PF00209; SNF; 1.
DR ProDom; PD000448; Na/ntran_sympot; 1.
DR PROSITE; PS0267; NA_NEUROTRAN_SYMP_3; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 241 AA; 27030 MW; AA15BDCBF5BBE07C CRC64;

Query Match 90.3%; Score 28; DB 11; Length 241;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
Db 188 LIHGKI 193

RESULT 41
Q8GSK7
ID Q8GSK7 PRELIMINARY; PRT; 252 AA.
AC Q8GSK7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P0438G07.22 protein (B1126F07.22 protein).
GN P0438G07.22 OR B1126F07.22.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
clone:P0438G07.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:B1126F07.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004398; BAC22375.1;
DR EMBL; AP005674; BAC22486.1;
SQ SEQUENCE 252 AA; 28503 MW; 0481F9D5201504BC CRC64;

Query Match 90.3%; Score 28; DB 10; Length 252;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6

```

```

Db      170 LLHGKL 175
RESULT 42
O9K7R5 PRELIMINARY; PRT; 254 AA.
ID O9K7R5
AC O9K7R5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3294.
GN BH3294.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
ON NCBI_TaxID=86665;
RX [1]
SEQUENCE FROM N.A.
STRAIN=C-125 / JCM 9153;
MDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07013.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 30129 MW; 6CB5E585051ECA4A5 CRC64;

Query Match 90.3%; Score 28; DB 16; Length 254;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 245 VVHGKL 250

RESULT 43
O9V4S8 PRELIMINARY; PRT; 278 AA.
ID O9V4S8
AC O9V4S8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG2038 protein.
GN CSN7B OR BCDNA.GH07822 OR CG2038.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balgwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Spierskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
SEQUENCE FROM N.A.
RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragass V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE FROM N.A.
RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003837; AAF59097.2; -
DR FlyBase; Fggn0033291; CSN7b.
DR InterPro; IPR000717; PCI.
DR Pfam; PF01399; PCI; 1.
DR SMART; SM00088; PINT; 1.
SQ SEQUENCE 278 AA; 31067 MW; CAB00E6C0E0A9BA8 CRC64;

Query Match 90.3%; Score 28; DB 5; Length 278;
Best Local Similarity 66.7%; Pred. No. 3.6e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVHGKL 6
Db 148 IHHGKL 153

```

RESULT 44

O04627 PRELIMINARY; PRT; 287 AA.
 AC O04627;
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE F2P16.1 protein.
 GN F2P16.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Miller N., Beck C., Kramer J., Bauer C.;
 RT "The sequence of A. thaliana F2P16.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA wasnu;
 RT "The A. thaliana Genome Sequencing Project.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Waterston R.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Wilson R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF007270; AAB61044.1; .
 DR InterPro; IPR001882; Biotin_attach.
 DR PROSITE; PS00188; BIOTIN; 1.
 SQ SEQUENCE 287 AA; 32657 MW; 8DB27EE87A750078 CRC64;

Query Match 90.3%; Score 28; DB 10; Length 287;
 Best Local Similarity 83.3%; Pred. No. 3.7e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 Db 67 LVHGRL 72

RESULT 45

Q8F7L2 PRELIMINARY; PRT; 312 AA.
 AC Q8F7L2;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Predicted hydrolases or acyltransferases alpha/beta hydrolase
 DE superfamily.
 GN LA0932.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
 RA Ren S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE011276; AAN48131.1; .
 KW Acyltransferase; Hydrolase; Transferase; Complete proteome.
 SQ SEQUENCE 312 AA; 36054 MW; 2A7F86BD85E7520A CRC64;

Query Match 90.3%; Score 28; DB 16; Length 312;
 Best Local Similarity 83.3%; Pred. No. 4.1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LVHGKL 6
 Db 190 LVHGKV 195

Search completed: October 14, 2003, 13:15:24
 Job time : 35.7273 secs

XX	23-JAN-1985;	85EP-0300432.
XX		
XX	24-JAN-1984;	84US-0573825.
PR	23-JUL-1987;	87US-0076811.
XX		
XX	(IMMV)	IMMUNEX CORP.
XX		
PI	Hopp TP,	Bektesh SL, Conlo
XX		
XX	WPI;	1985-185385/31.
XX		
PT	Synthesis of proteins with i	
PT	DNA techniques for economic	

PT chromatography

PS Disclosure; Page 24; 44pp; English.

XX The invention relates to a novel process for producing protein molecules
 CC by recombinant DNA techniques. The invention also relates to a novel DNA
 CC expression vector capable of expressing a heterologous polypeptide in a
 CC transformed host cell, the polypeptide consisting of a constituent
 CC polypeptide and an attached identification peptide. The identification
 CC peptide has an antigenic terminal portion of amino acids. Recombinant DNA
 CC techniques are used for the economic production of a desired protein. It
 CC is then efficiently purified in a single affinity chromatography step
 CC with high yields. The sequence represents a peptide from trypsinogen,
 CC which is the substrate for bovine mucosal enterokinase, and is used in
 CC the invention as a linking peptide in a hybrid molecule.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 6; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 1 DDDDK 5

RESULT 2

AAP81852

ID AAP81852 standard; protein; 5 AA.

XX AAP81852;

AC AAP81852;

XX 17-DEC-1990 (first entry)

XX Sequence encoded by synthetic adaptor linking human alpha-1-antitrypsin

DE (hAF) and human growth hormone releasing factor (hGRF).

XX Somatoliberin.

XX Synthetic.

XX Key Location/Qualifiers

FH Cleavage-site 5

XX EP252894-A.

XX 13-JAN-1988.

XX 13-MAY-1987; 87EP-0870070.

XX 15-MAY-1986; 86US-0863281.

XX (SMIK) SMITH KLINE - RIT.

XX Bollen AJ, Jacobs P;

XX WPI; 1988-008826/02.

XX N-PSDB; AAN81296.

XX New DNA coding for cleavable hybrid protein contg. somatocrocinin -
 PT linked to human alpha-anti-trypsin, providing efficient
 PT expression in bacteria

XX Claim 1; Fig 1; 10pp; French.

XX The synthetic adaptor codes for an AA SQ cleavable in vitro using
 CC enterokinase. hGRF is expressed much more efficiently in bacteria as a
 CC fusion protein, compared with cloning of hGRF gene alone.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 9; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 1 DDDDK 5

RESULT 3

AAR48070

ID AAR48070 standard; peptide; 5 AA.

XX AAR48070;

XX 25-MAR-2003 (updated)

DT 13-JUL-1994 (first entry)

XX Biological protecting gp. contg. enterokinase cleavage site.

XX C-terminal alpha-amide polypeptide; amidation; protecting group;
 KW N-terminal alpha-acetyl polypeptide; acetylation;
 KW recombinant multicopy fusion protein; interconnecting peptide;
 KW intraconnecting peptide; enterokinase; recognition site.

XX Synthetic.

XX Key Location/Qualifiers

FH Cleavage-site 1..5

FT /label= enterokinase_recognition_site

XX WO9401451-A2.

XX 20-JAN-1994.

XX 13-JUL-1993; 93WO-US06591.

XX 13-JUL-1992; 92US-0912798.

XX (BION-) BIONEBRASKA INC.

XX Coolidge TR, Holmquist B, Stout J, Wagner FW;

XX WPI; 1994-034983/04.

XX N-PSDB; AAQ55444.

XX Terminal modification of recombinant single copy polypeptide - by
 PT protecting, modifying and de-protecting polypeptide, e.g. fusion
 PT protein, contg. biologically added protecting gp.

XX Disclosure and Claim 10; Page 22; 102pp; English.

XX In order to produce recombinant polypeptides with C-terminal
 CC alpha-amide groups and/or N-terminal acetyl groups, fusion proteins
 CC are formed in which the relevant terminus (termini) of the
 CC polypeptide is protected by a biological protecting group. The
 CC protecting group is a peptide or amino acid having at least one
 CC cleavage site for its removal. The sequence AAR48070 represents a
 CC suitable biological protecting group, i.e. an enterokinase cleavage
 CC site.

XX (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 15; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 1 DDDDK 5

RESULT 4

AAR83117
 ID AAR83117 standard; Peptide; 5 AA.
 AC AAR83117;
 XX
 DT 29-APR-1996 (first entry)
 PR
 XX
 DE Enterokinase recognition sequence.
 XX
 KW Calcitonin; transgenic animal; milk; Paget disease; hypercalcaemia;
 KW osteoporosis; fusion protein cleavage; enterokinase; trypsinogen.
 XX
 OS Synthetic.
 XX
 PN WO9527782-A1.
 XX
 PD 19-OCT-1995.
 XX
 PF 05-APR-1995; 95WO-GB00769.
 XX
 PR 08-APR-1994; 94GB-0006974.
 XX
 PA (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.
 XX
 PI Cottingham IR, Garner I;
 XX
 DR WPI; 1995-366387/47.
 XX
 PT Peptide, pref. calcitonin, prodn. in transgenic non-human mammal -
 PT by expressing fusion protein in the milk followed by cleavage and
 PT sepn., used to treat Paget's disease, hypercalcaemic shock etc.
 XX
 PS Claim 14; Page 21; 26pp; English.
 XX
 CC The recognition sequence (AAR83117) for enterokinase, part of the
 CC activation peptide (AAR83118) of trypsinogen, may be included in
 CC the linker segment of a fusion protein, comprising e.g. calcitonin
 CC and alpha-lactalbumin, produced in the milk of a transgenic mammal.
 CC Cleavage of the fusion protein with enterokinase yields alpha-
 CC lactalbumin, which is removed by an affinity method, and purified
 CC calcitonin of therapeutic appln. Alternatively, the linker
 CC comprises the recognition sequence (AAR83116) of Factor-Xa.
 XX
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 16; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 DB 1 DDDDK 5
 RESULT 5
 AAW02129
 ID AAW02129 standard; Peptide; 5 AA.
 AC AAW02129;
 XX
 DT 29-OCT-1996 (first entry)
 PR
 DE Enterokinase cleavage site.
 XX
 KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KW src homology 2 domain; enterokinase.
 XX
 OS Synthetic.
 XX
 PN EP727211-A1.
 XX
 PD 21-AUG-1996.
 XX

PF 07-FEB-1996; 96EP-0200270.
 XX
 PR 29-DEC-1995; 95US-0580868.
 PR 10-FEB-1995; 95US-0386381.
 PR 07-MAR-1995; 95US-0400220.
 PR 30-JUN-1995; 95US-0497357.
 PR 11-OCT-1995; 95US-0541080.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Dunnington DJ;
 XX
 DR WPI; 1996-372674/38.
 XX
 PT Use of selective src SH2 domain ligand - to prepare medicament for
 PT treating bone resorption disease
 XX
 PS Example 11; Page 27; 47pp; English.
 XX
 CC Tagged protein DET1-DET2-spacer-ek-SH2 comprises: DET1 (AAW02128), or
 CC defined epitope tag 1, a peptide found in HIV-1 gp120/gp160, an
 CC epitope for monoclonal antibody 178.1; DET2, or defined epitope tag
 CC 2, a hexahistidine tag which facilitates purification; a spacer
 CC utilised to design a BamHI site; ek (AAW02129), a recognition sequence
 CC for enterokinase; and a protein SH2 domain. Tagged proteins
 CC including SH2 domains from human src, lck and hcp (see also AAW02119-21)
 CC were utilised in binding assays to determine the specificity of cpds.
 CC to inhibit SH2 domains; cpds. that selectively inhibit the human src
 CC SH2 domain can be used to treat bone resorption diseases such as
 CC osteoporosis.
 XX
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 17; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 DB 1 DDDDK 5
 RESULT 6
 AAW32849
 ID AAW32849 standard; Peptide; 5 AA.
 AC AAW32849;
 XX
 DT 25-MAR-2003 (updated)
 DT 15-JAN-1998 (first entry)
 XX
 DE Protecting group with enterokinase cleavage site.
 XX
 KW Preparation; modified recombinantly produced polypeptide;
 KW biologically added protecting group; modification;
 KW enterokinase.
 XX
 OS Synthetic.
 XX
 PN US5656456-A.
 XX
 PD 12-AUG-1997.
 XX
 PF 01-JUN-1995; 95US-0457166.
 XX
 PR 13-JUL-1993; 93US-0091751.
 PR 14-JUL-1992; 92US-0912798.
 PR 01-JUN-1995; 95US-0457166.
 XX
 PA (BION-) BIONEERASKA INC.
 XX
 PI Coolidge TR, Holmquist B, Stout J, Wagner FW;
 XX

DR WPI; 1997-414586/38.
 XX N-PSDB; AAT88174.
 XX
 PT Selective modification of recombinantly produced polypeptide - at
 PT either or both terminals, by reaction with chemical protecting
 PT agent, cleaving biologically added protecting group and reacting
 PT with chemical modifying agent
 XX
 PS Disclosure; Columns 43-44; 28pp; English.
 XX
 CC Preparing a modified recombinantly produced polypeptide, comprising
 CC a reactive side chain group including an NH₂, OH, SH or COOH group,
 CC and a modified amino-terminal alpha-amino (NAN) group or a
 CC carboxy-terminal alpha-carboxy (CAC) group, comprises recombinantly
 CC forming a polypeptide having a biologically added protecting group
 CC (BAPG), i.e. the present BAPG, on an NAN or CAC group. Then the
 CC reactive side chain group is reacted with a chemical protecting
 CC agent and the BAPG cleaved with a cleavage reagent to form the
 CC unprotected terminal amino acid alpha-carbon reactive group, which
 CC is modified with a chemical modifying agent to form a side chain
 CC protected polypeptide.
 CC The method is useful for the modification of naturally occurring
 CC proteins and peptides, e.g. to change the structural
 CC characteristics and/or biological activity, provide targeting to a
 CC specific cell type, change the rate of breakdown and clearance,
 CC increase the biological activity/potency or form antagonists.
 CC Typically amino-amidation of small peptides (e.g. mastoparan or
 CC human gastrin releasing peptide) enhances the biological activity.
 CC The NAN and/or CAC groups of a recombinantly produced polypeptide
 CC can be selectively modified by a wide range of organic groups,
 CC without adversely affecting the reactive side-chain groups. The
 CC method is convenient, cheap and gives high yields.
 CC (Updated on 23-MAR-2003 to correct PF field.)
 XX
 XX Sequence 5 AA;
 SQ

Query Match 100.0%; Score 29; DB 18; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 1 DDDDK 5

RESULT 7
 AAW19622
 ID AAW19622 standard; peptide; 5 AA.
 AC AAW19622;
 XX
 XX 21-OCT-1997 (first entry)
 DE Enterokinase protease recognition site.
 XX
 DE Stat 5: Signal Transduction and Activation of Transcription;
 KW Src homology domain; SH2; erythropoiesis enhancing; anaemia;
 KW fusion protein; ek; enterokinase; epitope; antibody production;
 KW detection; HIV; human immunodeficiency virus type 1; gp120;
 KW glycoprotein 120; selective.
 XX
 OS Homo sapiens.
 XX
 PN WO9702024-A1.
 XX
 PD 23-JAN-1997.
 XX
 PF 28-JUN-1996; 96WO-US11158.
 XX
 PR 08-FEB-1996; 96US-0598715.
 PR 30-JUN-1995; 95US-0497357.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.

XX Dunnington DJ;
 PI WPI; 1997-108736/10.
 DR
 XX
 PT Enhancing erythropoiesis with specific activator of human Stat 5 SH2
 PT domain - has very low binding affinity to other SH2 domains so free
 PT of side effects, particularly for treating anaemia
 XX
 PS Example 11; Page 53; 91pp; English.
 XX
 CC AAW19622 is an enterokinase protease recognition site (ek) this
 CC sequence is included in a fusion protein to provide a site for cleavage
 CC of the fusion protein into its constituent parts. Fusion proteins (FP)
 CC of the Stat5 Src homology domain SH2 were constructed. FP had the
 CC formula DFT1-DFT3-Sp-ek-SH2, where Sp is a spacer. DFT is a defined
 CC epitope tag and SH2 is the Stat5 SH2 domain. The DFTs are included so
 CC that antibodies against these epitopes can be used to detect the
 CC recombinant expression of the fusion protein and hence the SH2 domain.
 CC The fusion proteins are used for identifying compounds that bind the
 CC SH2 domain causing its activation, mimicking erythropoietin. Compounds
 CC identified will preferably bind to human Stat5 SH2 domain with an
 CC affinity between 50-100 times the binding affinity of human Stat6,
 CC hcp, SH-PP2, p85, grb2, src, lck or fyn SH2 domains. Compounds
 CC identified are useful for the treatment of anaemia and avoid the side
 CC effects (e.g. exacerbation of allergic reactions) associated with
 CC non-selective inhibition of SH2 domains.
 XX
 SQ Sequence 5 AA;
 SQ

Query Match 100.0%; Score 29; DB 18; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 1 DDDDK 5

RESULT 8
 AAW21758
 ID AAW21758 standard; peptide; 5 AA.
 XX
 AC AAW21758;
 XX
 XX 27-AUG-1997 (first entry)
 DE Enterokinase cleavage site linker for globin fusion protein.
 XX
 DE Cleavable linker; haemoglobin; alpha globin; beta globin; magainin;
 KW hirudin; alpha-endorphin; dynorphin; calcitonin; oxytocin;
 KW fusion protein; recombinant protein production; enterokinase.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "optionally absent"
 FT
 PN US5627268-A.
 XX
 PD 06-MAY-1997.
 XX
 PF 07-JUN-1994; 94US-0255272.
 XX
 PR 07-JUN-1994; 94US-0255272.
 XX
 PA (DNXB-) DNX BIOTHERAPEUTICS.
 XX
 PI Khoury-Christianson AM, Kumar R, Sharma A;
 DR WPI; 1997-271367/24.
 XX

PT Haemoglobin containing fusion protein - comprising globin
 XX polypeptide and desired peptide

XX Claim 5; Column 32; 47pp; English.

XX Fusion polypeptides have been produced in which a globin polypeptide
 CC is attached to a polypeptide having fewer than 65 amino acids
 CC (especially magainin, hirudin, an endorphin, a dynorphin, calcitonin
 CC or oxytocin) through a selectively cleavable peptide bond. The net
 CC charge of the selectively cleavable peptide bond and the desired
 CC protein is -4 to +5 to allow fusion proteins to be successfully
 CC assembled into tetrameric haemoglobin. Such chimeric haemoglobin can
 CC be produced in large quantities in erythrocytes of transgenic
 CC animals and the desired peptide can be obtained from a haemolysate
 CC of the red blood cells by enzymatic or chemical cleavage at the linker.
 CC The present sequence is a preferred linker peptide suitable for use
 CC in the new fusion polypeptides.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 18; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 1 DDDDK 5

RESULT 9

AAW75805
 ID AAW75805 standard; peptide; 5 AA.

XX AC AAW75805;

XX DT 27-OCT-1998 (first entry)

XX DE Enterokinase susceptible peptide sequence.

XX Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;
 KW trypsin-6 protein; inflammatory disorder; allergic rhinitis; urticaria;
 KW antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;
 KW hyperproliferative skin disease; peptic ulcer; hyperresponsiveness;
 KW inflammatory skin condition; enterokinase.

XX OS Unidentified.

XX PN WO9833812-A1.

XX PD 06-AUG-1998.

XX PF 30-JAN-1998; 98WO-US01865.

XX PR 05-FEB-1997; 97US-0037090.

XX PS (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX PI Huang C, Stevens RL;

XX PS WPI; 1998-437390/37.

XX Trypsin-6 complex inhibitory peptides - used to treat mast
 PT cell-mediated inflammatory disorders e.g. asthma

XX PS Example; Page 18; 69pp; English.

XX This is an enterokinase susceptible peptide sequence. The invention
 CC provides sequences shown in AAW63160 to AAW63169 that are inhibitors of
 CC mouse mast cell protease (mMCP-6). These peptides which are tryptase-6
 CC complex inhibitors, can be used for treating a mast cell-mediated
 CC inflammatory disorder. The inhibitors can be used to treat inflammatory
 CC disorders including asthma, allergic rhinitis, urticaria and antioedema,
 CC eczematous dermatitis (atopic dermatitis), anaphylaxis,

CC hyperproliferative skin disease, peptic ulcers, inflammatory bowel
 CC disorder, hyperresponsiveness and inflammatory skin conditions.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 19; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 1 DDDDK 5

RESULT 10

AAAY50206
 ID AAAY50206 standard; Peptide; 5 AA.

XX AC AAAY50206;

XX DT 12-JAN-2000 (first entry)

XX DE Neutrophil-activating pancreatic derived peptide 6.

XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;
 KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;
 KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;
 KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;
 KW trauma; protease inhibitor; hypertension; sepsis.

XX OS Cricetulus sp.

XX PN WO9946367-A2.

XX PD 16-SEP-1999.

XX PF 11-MAR-1999; 99WO-US05247.

XX PR 11-MAR-1998; 98US-0038894.

XX PA (CELL-) CELL ACTIVATION INC.

XX PA (REGC) UNIV CALIFORNIA.

XX PA (SCRI) SCRIPPS RES INST.

XX PI Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;

XX PS WPI; 1999-580234/49.

XX Use of cell activating compositions in developing products for
 CC diagnosis and treatment of e.g. cardiovascular, inflammatory,
 CC autoimmune or Alzheimer's disease, trauma, arthritis, organ rejection,
 CC diabetes, stroke or ischemia -

XX PS Example 9; Page 182; 184pp; English.

XX This invention describes a novel method for the use and preparation of
 CC cell activating compositions which involves preparing a cell activating
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at
 CC about neutral or higher pH to produce a homogenate; (b) removing the
 CC particulates from the homogenate; (c) optionally incubating the
 CC resulting homogenate, with particulates removed, with a protease; and
 CC (d) fractionating the homogenate and selecting fractions that exhibit
 CC cell activation activity. The methods can be used for improving
 CC treatment outcome or reducing risk of treatment of e.g. cardiovascular
 CC disease, inflammatory disease, trauma, autoimmune diseases, arthritis,
 CC organ rejection, diabetes and diabetic complications, stroke, ischemia,
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.
 CC They can be used in the veterinary treatment of a non-human subject.
 CC Protease inhibitors can be used to lower cell activation resulting from
 CC these diseases and deficiencies. The detection of an elevated level of
 CC hydrogen peroxide can be used to detect an inflammatory condition. An
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the

CC presence of superoxide dismutase (SOD) indicates leukocyte up
 CC regulation, e.g. indicative of the onset of an acute cardiovascular
 CC disorders, such as disease onset or ischemic complications. An elevated
 CC level of hydrogen peroxide in plasma or whole blood and a low level in
 CC the presence of SOD is indicative of a chronic or immune compromised
 CC condition e.g. hypertension or sepsis. AAY50201-Y50334 represent peptides
 CC used in the method of the invention.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 20; Length 5;
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
 |||||
 Db 1 DDDDK 5

RESULT 11

AAO20103
 ID AAO20103 standard; Peptide; 5 AA.

AC AAO20103;

DT 06-AUG-2002 (first entry)

DE Alpha-human atrial natriuretic factor related peptide #2.

KW Microbiological preparation; alpha-human atrial natriuretic factor.

XX Unidentified.

XX KR99081178-A.

XX 15-NOV-1999.

XX 27-APR-1998; 98KR-0014975.

XX 27-APR-1998; 98KR-0014975.

XX (CHEI-) CHEIL FOODS & CHEM INC.

XX Oh MS, Koh HG, Lee DU, Hah BJ, Park JS;

XX WPI; 2000-617253/59.

XX N-PSDB; AAK99521.

XX Microbiological preparation method of alpha-human atrial natriuretic
 factor -

XX Disclosure; Page 3; 10pp; Korean.

XX The invention relates to a microbiological preparation method of alpha-
 CC human atrial natriuretic factor. This sequence represents a peptide
 CC sequence relating to the alpha-human atrial natriuretic factor of the
 CC invention.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
 |||||
 Db 1 DDDDK 5

RESULT 12

AAB35703
 ID AAB35703 standard; Protein; 5 AA.

XX

AC AAB35703;

DT 14-FEB-2001 (first entry)

DE Peptide 2 used in trypsin hL identification.

XX Human; trypsin hL; serine protease; lung disease model animal.

XX Synthetic.

PN JP2000253887-A.

XX 19-SEP-2000.

XX 11-MAR-1999; 99JP-0065337.

XX 11-MAR-1999; 99JP-0065337.

XX (TTPH-) TT PHARMA KK.

XX WPI; 2000-658814/64.

XX Novel gene encoding a serine protease and its protein used to screen
 PT for serine protease inhibitors and to prepare lung disease animal
 PT models -

XX Disclosure; Page 9; 17pp; Japanese.

XX Nucleotide sequence AAC66182 encodes human trypsin hL AAB35701, a serine
 CC protease. The invention relates to the human hL gene and protein
 CC sequences, to a recombinant vector containing the nucleotide sequence,
 CC and a host cell containing the vector. Human trypsin hL can be used for
 CC screening for serine protease inhibitors, in the preparation of a lung
 CC disease model animal, and for the development of an index marker of lung
 CC diseases caused by an anti-trypsin hL antibody. The present sequence
 CC represents a peptide used in the identification of trypsin hL.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
 |||||
 Db 1 DDDDK 5

RESULT 13

AAB15680
 ID AAB15680 standard; peptide; 5 AA.

AC AAB15680;

DT 08-JAN-2001 (first entry)

DE Enterokinase-sensitive linker peptide.

XX Enterokinase; restriction protease; protease-sensitive linker;
 KW stabilizing fusion peptide; single-chain antibody;
 KW vaccine; gene therapy; protein degradation modulation;
 KW protein stability; Alzheimer's disease.

XX Unidentified.

XX WO2000042185-A1.

XX 20-JUL-2000.

XX 11-JAN-2000; 2000WO-US00558.

XX 11-JAN-1999; 99US-0115505.

XX

PA (MIND-) MINDSET BIOPHARMACEUTICALS USA INC.
PA (MCIN/) MCINNIS P A.

Chain DG;

DR WPI; 2000-476059/41.

XX Modulating degradability of protein or peptide useful for gene therapy
PT involving antibodies, comprises altering a gene at the N terminus to
PT render protein or peptide metabolically stable -

XX Disclosure; Page 12; 76pp; English.

XX The present sequence is a peptide which is sensitive to the restriction
CC protease enterokinase. It may be used to link a stabilising fusion
CC peptide, referred to as a stabilon, to the N-terminus of a single-chain
CC antibody. The stabilon increases stability of the antibody against
CC proteolysis in vivo. Induction of expression of enterokinase by means of
CC an inducible promoter causes cleavage of the linker and removal of the
CC stabilon. This renders the antibody susceptible to proteolysis by the
CC N-end rule pathway. Attachment of the stabilon to the antibody through a
CC protease-sensitive linker region therefore allows modulation of antibody
CC stability. Stabilised recombinant proteins may be used in gene therapy
CC for the treatment of disorders such as Alzheimer's disease.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 14

AAB22833

ID AAB22833 standard; peptide; 5 AA.

AC AAB22833;

XX 10-JAN-2001 (first entry)

DE Enterokinase cleavage site.

XX Controlled release delivery system; drug targetting;
KW drug-specific antibody; intracellular half-life; gene therapy;
KW diabetes; autoimmune disease; inflammatory disease; infectious disease;
KW cancer; side effect; stabilon; protease cleavage site.

XX Unidentified.

OS WO200050089-A2.

PN 31-AUG-2000.

XX 25-FEB-2000; 2000WO-US04749.

XX 26-FEB-1999; 99US-0122103.

XX (MIND-) MINDSET BIOPHARMACEUTICALS USA INC.

Chain DG;

DR WPI; 2000-572044/53.

XX Delivering a drug, to a patient suffering from cancer or diabetes, at a
PT predetermined site, comprises altering a target cell to express a drug
PT specific antibody by gene therapy and administering the drug -

PS Disclosure; Page 21; 69pp; English.

XX

CC The invention relates to a controlled release drug delivery system.
CC Target cells in a patient are transformed with an expression construct
CC encoding an antibody (particularly a single chain antibody) to the drug
CC to be delivered. The antibody contains a modulator of intracellular
CC half-life; this can either be a stabilising or destabilising residue
CC located in the N-terminus after the initial methionine, or a peptide
CC ("stabilon") containing a stabilising residue linked to the antibody
CC N-terminus via a protease cleavage site. On administration of the drug,
CC the antibody binds the drug, localising it at the target cells but
CC maintaining it in an inactive state. As the antibody is degraded (the
CC timescale for which is dependent upon the N-terminal or stabilon amino
CC acids), the drug is released at its site of action where it can exert
CC its effects. The antibody encoded by the expression construct (and
CC therefore the therapeutic agent) can be targeted to particular
CC subcellular locations (e.g., the nucleus) by including the appropriate
CC cellular localisation signals. The novel method may be used to deliver
CC therapeutic agents to patients with a variety of conditions such as
CC diabetes, autoimmune diseases, inflammatory diseases, infectious
CC diseases and especially cancer. The delivery method of the invention
CC causes a drug to become almost immediately localised at its site of
CC action in an inactive form where it accumulates. Once released by
CC antibody degradation, the drug is at an effective concentration only at
CC the target site, with very little free drug being available in the rest
CC of the body. The system of the invention therefore reduces the side
CC effects caused by therapeutic agents, and also provides economic benefits
CC as a smaller amount can be administered to the patient. Sequences
CC AAB22832- AAB22834 represent protease cleavage sites which can be used to
CC link a stabilon to the N-terminus of a drug-specific antibody used in
CC the method of the invention.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 15

AAB23830

ID AAB23830 standard; Peptide; 5 AA.

XX AAB23830;

XX 15-JAN-2001 (first entry)

DE Linker peptide recognised by enterokinase.

XX Human; glucagon; fusion expression partner; interleukin 2; IL-2;
KW linker.

XX Synthetic.

OS WO200053777-A1.

PN 14-SEP-2000.

XX 09-MAR-2000; 2000WO-KR00187.

XX 09-MAR-1999; 99KR-0007641.

XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

PA (HANW-) HANWHA CHEM CORP.

XX Park YH, Lee J, Kim D;

XX WPI; 2000-611446/58.

XX Preparation of recombinant proteins in the form of fusion proteins
PT having human glucagon sequence as a fusion expression partner useful

PT for preparing human interleukin-2 or its derivatives in microorganisms

XX

PS Claim 5; Page 35; 51pp; English.

XX

CC The present invention describes the preparation of recombinant proteins

CC (1) as fusion proteins having human glucagon sequence as a fusion

CC expression partner. The method is useful for preparing human

CC interleukin-2 (IL-2) or its derivatives in microorganisms. The method

CC comprises culturing *Escherichia coli* transformants producing human

CC IL-2 or its derivatives fused with fusion partners, obtaining protein

CC aggregates from the cell lysates and solubilising the protein aggregates

CC in alkaline solution and renaturing recombinant protein by restoring

CC neutral pH. The fusion protein is digested with enterokinase during

CC purification. The preparation method is useful for maximising the

CC production yield of recombinant proteins in cells in the form of

CC soluble inclusion bodies, which are soluble in alkaline solution

CC without using denaturants and reducing agents. The present sequence

CC represents a specifically claimed peptide linker which is recognised

CC by enterokinase.

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDDDK 5

Db 1 DDDDK 5

|||||

RESULT 16

AAB12953

ID AAB12953 standard; peptide; 5 AA.

AC AAB12953;

XX

DT 23-NOV-2000 (first entry)

DE Enterokinase recognition and cleavage site.

XX

KW Streptokinase; platelet glycoprotein binding peptide; clot dissolution;

KW thromboembolic disorder; thrombolytic; vasotropic; pulmonary embolism;

KW cerebroprotective; ischaemic stroke; myocardial infarction; cardiant;

KW enterokinase recognition site.

XX

OS Synthetic.

XX

PN US6087332-A.

XX

PD 11-JUL-2000.

XX

PF 23-DEC-1997; 97US-0997532.

XX

PR 23-DEC-1997; 97US-0997532.

XX

PA (GALL/) GALLER L I.

XX

PI Galler LI;

XX

DR WPI; 2000-490194/43.

XX

XX New streptokinase derivatives with platelet glycoprotein binding

PT domains adjacent to the termini of the streptokinase sequence, useful

PT for treating thromboembolic disorders, e.g. ischemic stroke or massive

PT pulmonary embolism

XX

PS Disclosure; Column 5; 15pp; English.

XX

CC This invention relates to a recombinant DNA sequence comprising DNA

CC encoding a streptokinase, a polylinker or restriction sequence, and a

CC DNA sequence encoding a platelet glycoprotein binding peptide. The

CC protein encoded by the recombinant DNA of the invention is used in a

CC method for the treatment of thromboembolic disorders. Streptokinase forms

CC a 1:1 complex with human plasminogen and ultimately aids in the formation

CC of plasmin which is a fibrinolytic enzyme. Sufficient degradation of

CC fibrin elements in a clot results in clot dissolution. The recombinant

CC DNA of the invention has thrombolytic, vasotropic, cardiant, and

CC cerebroprotective activity, and works as a platelet aggregation

CC inhibitor. The compounds or the invention are useful for treating

CC thromboembolic disorders such as ischaemic stroke, myocardial infarction,

CC or massive pulmonary embolism.

CC This sequence represents an enterokinase recognition and cleavage site.

CC This cleavage site is contained in a plasmid used in the invention for

CC the expression of the streptokinase derivative fusion proteins encoded by

CC the recombinant DNA.

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDDDK 5

Db 1 DDDDK 5

|||||

RESULT 17

AAB12387

ID AAB12387 standard; Peptide; 5 AA.

AC AAB12387;

DT 06-NOV-2000 (first entry)

DE Enterokinase cut site.

XX GF-14; 14-3-3; solubility; fusion protein; enterokinase.

XX Unidentified.

XX US6077689-A.

XX 20-JUN-2000.

PD 24-DEC-1997; 97US-0997918.

PF 24-DEC-1997; 97US-0997918.

PR 24-DEC-1997; 97US-0997918.

XX (AMGE-) AMGEN INC.

XX Shavely MD;

XX WPI; 2000-450937/39.

XX Increasing the solubility of a protein produced in a host cell, useful

PT for enhancing direct expression of recombinantly produced polypeptides,

PT by expressing the recombinant protein as a fusion protein with a 14-3-3

PT protein, e.g. GF-14

XX Example 3; Column 6; 32pp; English.

XX The present invention relates to a method of increasing the solubility

CC of a protein. The method involves creating a fusion protein with a

CC 14-3-3 protein, for example Arabidopsis GF-14. The fusion protein can

CC be expressed in a host cell such as *E. coli*. An artificial form of

CC GF-14 DNA was created in which the codons were optimised for expression

CC in *E. coli*. This modified sequence was made by joining a series of

CC oligonucleotides. This method enhances the solubility of recombinant

CC proteins by reducing inclusion body formation. This allows molecular

CC chaperones more time in which to interact with a slowly-folding,

CC rapidly-produced, aggregation-prone protein. Costly and time consuming

CC refolding procedures are also avoided. The present sequence is the

CC enterokinase cut site. This sequence may be included in the linker

CC peptide between the two proteins of the fusion. This allows the GP-14
CC protein to be removed after the fusion protein has been generated.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 18

AA96953
ID AAY96953 standard; peptide; 5 AA.

XX AAY96953;

AC AAY96953;

DT 31-OCT-2000 (first entry)

XX Enterokinase site for use in rubredoxin fusion protein.

DE PRUBEX3; polyhistidine; tag; rubredoxin; antigen; vaccine; carrier;

KW immunogenic; enterokinase; flag tag; fusion.

XX Synthetic.

OS WO200039310-A1.

XX 06-JUL-2000.

XX 29-DEC-1999; 99WO-US31176.

XX 29-DEC-1998; 98US-0114034.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Przybyla A, Menon N;

XX WPI; 2000-452403/39.

XX Nucleic acids encoding fusion peptides comprising rubredoxin, useful as
PT vaccines

PS Disclosure; Page 62; 67pp; English.

CC Fusion proteins comprising rubredoxin and fused antigens and/or haptens
CC may be used as vaccines to initiate immune responses. In this case, the
CC rubredoxin acts as a carrier. The fusion protein is capable of binding
CC iron (Fe-2+) when properly folded, giving it a red color that makes it
CC easy to identify following or during purification. The C-terminal fused
CC protein may be insoluble or known to form inclusion bodies in a host
CC cell. The rubredoxin serves as the carrier molecule to yield immunogenic
CC fusion products. As rubredoxin is itself only negligibly antigenic, there
CC is no need to include a cleavage site in the fusion product to allow
CC cleavage of the N-terminal and C-terminal constituents.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 19

AA03861

ID AAB03861 standard; peptide; 5 AA.

XX AAB03861;

AC 20-OCT-2000 (first entry)

DE Peptide #2 used in expression vector construction.

XX Expression vector; active target protein; easy isolation.

XX Unidentified.

OS WO200031284-A1.

PN 02-JUN-2000.

XX 19-NOV-1999; 99WO-JP06474.

XX 20-NOV-1998; 98JP-0331515.

XX (FUSO) FUSO PHARM IND LTD.

XX Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;

XX WPI; 2000-411958/35.

XX Protein expression vectors containing specific sequences with cloning
PT sites into which nucleic acids insert, for isolation of target proteins
PT without excess amino acids -

PS Claim 8; Page 21; 44pp; Japanese.

XX The invention relates to a protein expression vector from which mature
CC and active target proteins can be easily isolated. N and C-terminal amino
CC acids are kept to a minimum to ensure that easy isolation from the vector
CC is possible. The protein expression vector comprises a secretory signal
CC nucleic acid sequence and, in the 3'-downstream side, a tag nucleic acid
CC sequence, a scissile nucleic acid sequence and a cloning site into which
CC a nucleic acid sequence encoding a target protein can be inserted. Also
CC included in the invention are a host cell transformed by the vector, a
CC process for producing a target protein using the vector or host cell, a
CC recombinant fusion protein containing the amino acid sequence of the
CC target protein, and a process for producing a target protein. The
CC expression vector is used for the production of mature and active target
CC protein. The present sequence represents a peptide used in the
CC construction of the vector of the invention.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 21; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 20

AA90830

ID AAY90830 standard; Peptide; 5 AA.

XX AAY90830;

XX 29-AUG-2000 (first entry)

DE Enterokinase peptide sequence SEQ ID NO:33.

XX Antigen binding site; immunoglobulin; cancer antigen; immunological;
KW antibody; tumour; human; mucin; cancer; cytostatic; hybridoma;
KW specific binding assay; affinity purification; drug targeting;
KW toxin targeting; imaging; genetic; therapeutic.

XX Homo sapiens.

OS

XX US6054561-A.
 XX 25-APR-2000.
 XX 07-JUN-1995; 95US-0483749.
 XX 21-MAR-1986; 86US-0842476.
 XX 08-MAY-1988; 88US-0190778.
 XX 08-FEB-1984; 84US-0577976.
 XX 11-JAN-1985; 85US-0690750.
 XX 11-AUG-1994; 94US-0288981.
 XX (CHIR) CHIRON CORP.
 XX Ring DB;
 XX WPI; 2000-338508/29.
 XX Monoclonal antibody capable of binding to human breast cancer antigen
 useful for affinity purification, drug or toxin targeting, imaging, and
 treating cancer
 XX Disclosure; Column 47; 57pp; English.
 XX The present invention describes a monoclonal antibody (MAB) (I) that
 binds to a human breast cancer antigen that is also bound by MAb 454C11
 and 520C9 (produced hybridoma ATCC HB8484 and HB8696, respectively). Also
 described is a hybridoma that produces (I). (I) is useful in specific
 binding assays, affinity purification, drug or toxin targeting, imaging,
 and genetic or immunological therapeutics for various cancers. The
 present sequence represents a peptide amino acid sequence which is used
 in the exemplification of the present invention.
 XX Sequence 5 AA;
 SQ Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 RESULT 21
 AAY90452
 ID AAY90452 standard; peptide; 5 AA.
 XX AC AAY90452;
 XX 18-JUL-2000 (first entry)
 XX Enterokinase substrate linker, SEQ ID NO:70.
 XX Targetted gene delivery; fibroblast growth factor receptor;
 KW FGFR-binding protein; nucleic acid binding protein;
 KW receptor-internalised ligand; cytotoxin; saporin; gene therapy;
 KW cytochrome; antiproliferative; cancer; melanoma; diabetic retinopathy;
 KW rheumatoid arthritis; restenosis, Dupuytren's contracture; psoriasis;
 KW eczema; nuclear translocation signal; NTS;
 KW cytoplasmic translocation signal; endosome-disruptive peptide.
 XX Unidentified.
 XX OS
 XX US6037329-A.
 XX 14-MAR-2000.
 XX 24-SEP-1996; 96US-0718904.
 XX 15-MAR-1994; 94US-0213446.
 XX 15-MAR-1994; 94US-0213447.

PR 29-AUG-1994; 94US-0297961.
 PR 13-SEP-1994; 94US-0305771.
 PR 16-MAY-1995; 95US-0441979.
 XX (SELE-) SELECTIVE GENETICS INC.
 XX Chandler LA, Sosnowski BA, Baird JA;
 PI WPI; 2000-292008/25.
 DR N-PSDB; AAA12886.
 XX Gene delivery system, useful for treating or preventing cancer and
 rheumatoid arthritis, comprises receptor-internalized ligand linked to
 nucleic acid binding domain and nucleic acid
 XX Example 7; Column 69; 131pp; English.
 XX The invention relates to a novel gene delivery composition for the
 targetted delivery of cytotoxins or prodrug-converting enzymes to
 proliferating cells. The gene delivery composition comprises a protein
 that binds the fibroblast growth factor receptor (FGFR) which is fused
 or chemically conjugated to a nucleic acid binding domain. The nucleic
 acid binding domain is complexed with a suitable expression construct
 encoding a cytotoxin such as saporin. One or more linkers may join the
 FGFR-binding protein to the nucleic acid binding protein. These are
 selected to increase the specificity, toxicity, solubility, serum
 stability or intracellular availability, and may serve to promote
 condensation of nucleic acids for delivery to a cell. The fusion protein
 binds to FGFR and is internalised by cells that carry this receptor. The
 gene delivery composition is used for the therapeutic alteration of the
 function, gene expression and viability of cells. In particular, it may
 be used for the treatment and prevention of cell proliferative
 disorders, for example after eye surgery, melanoma and many other sorts
 of cancer, rheumatoid arthritis, restenosis, Dupuytren's contracture,
 diabetic retinopathy, psoriasis and eczema. The gene delivery
 compositions of the invention have high specificity for particular cells
 and can deliver larger amounts of DNA compared to prior art methods.
 CC Sequences AAY90424- AAY90441 represent nuclear translocation signals
 (NTSS) from a variety of proteins, and AAY90442-Y90444 represent
 cytoplasmic translocation signals. Sequences AAY90445-Y90446 are
 endosome-disruptive peptides. Sequences AAA12863-AL2867 encode flexible
 linker peptides, AAY90447 being a claimed linker peptide. AAA12882-AL2887
 CC encode linkers that are substrates for certain enzymes. AAA12882-AL2887
 CC being examples of such linkers.
 XX SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 RESULT 22
 AAY44370
 ID AAY44370 standard; peptide; 5 AA.
 XX AC AAY44370;
 XX 14-MAR-2000 (first entry)
 XX Enterokinase recognition peptide.
 XX Proinsulin; cleavage site; protease; retroviral vector;
 KW Lhr asterisk TgruSN; glucose-regulated protease; TGF alpha;
 KW transforming growth factor alpha promoter; selectable marker;
 KW neomycin phosphotransferase; insulin; diabetes mellitus;
 KW enterokinase recognition peptide.
 XX Homo sapiens.

XX PN WO9963101-A2.
 XX PD 09-DEC-1999.
 XX PF 01-JUN-1999; 99WO-US11970.
 XX PR 02-JUN-1998; 98US-0087660.
 XX PR 04-NOV-1998; 98US-0185852.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Osborne WRA, Ramesh N;
 XX DR WPI; 2000-105702/09.
 XX PT Novel population of transduced cells for treating/preventing diabetes
 XX PT and its related pathological conditions -
 XX PS Disclosure; Page 37; 104pp; English.
 XX CC The present sequence is a pentapeptide that can be recognised
 CC and cleaved after Lysine by protease, enterokinase. Human proinsulin
 CC cleavage site can be modified by introducing the recognition peptide for
 CC cleavage by protease and the cDNA can be inserted into a
 CC three-gene retroviral vector Lhi asterisk TgFUSN. The vector also
 CC comprises nucleic acids encoding glucose-regulated protease, glucose-
 CC responsive transforming growth factor (TGF) alpha promoter and selectable
 CC marker neomycin phosphotransferase. This vector can be introduced into
 CC smooth muscle cells for expression of bioactive insulin that can be used
 CC for treating diabetes. These cells can be used for diagnosis, treatment
 CC and prevention of diabetes mellitus and can also be used to improve the
 CC efficacy of other therapies for diabetes.
 XX SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 RESULT 23
 AAY57592
 ID AAY57592 standard; peptide; 5 AA.
 AC AAY57592;
 XX 02-MAR-2000 (first entry)
 DE Enterokinase recognition and cleavage site.
 XX Human; heparanase; hpa; genetic modification; expression; anticancer;
 KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
 KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
 KW micrometastasis; autoimmune lesion; kidney failure.
 XX Synthetic.
 OS WO9957244-A1.
 XX 11-NOV-1999.
 XX 29-APR-1999; 99WO-US09256.
 XX 01-MAY-1998; 98US-0071618.
 XX 02-MAR-1999; 99US-0260038.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX (FRIE/) FRIEDMAN M M.
 XX PI Ben-Artzi H, Ayal-HersHKovitz M, Yacoby-Zeevi O, Pecker I, Peleg Y;
 XX PI Shlomi Y;
 XX DR WPI; 2000-062144/05.
 XX PT Engineered cells that express recombinant heparanase, useful
 XX PT therapeutically, e.g. for treating angiogenesis and to screen for
 XX PT specific inhibitors, potential anticancer agents -
 XX PS Example 6; Page 54; 118pp; English.
 XX CC The present invention describes genetically modified cells (A) containing
 CC a polynucleotide (i) that encodes a polypeptide with heparanase activity,
 CC and express recombinant heparanase (ii). Heparanase cleaves heparan
 CC sulphate (HS) at specific intrachain sites, resulting in release of
 CC heparin-binding growth factors, enzymes and proteins that are sequestered
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It
 CC may also be implicated in tumour angiogenesis and metastases. (ii) is
 CC potentially useful in wound healing and for treating angiogenesis,
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
 CC infection and cystic fibrosis. It can also be used to neutralise heparin
 CC (an alternative to protamine) and to screen for specific inhibitors
 CC (potentially useful for treating cancer and metastases). Antibodies
 CC raised against (ii) are used for immunodetection and diagnosis of
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (ii)
 CC in large quantities, in a form that is homogeneously processed and
 CC activated/neutralised by a dedicated protease. The present sequence
 CC represents a peptide from an example from the present invention.
 XX SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 RESULT 24
 ABB08508
 ID ABB08508 standard; peptide; 5 AA.
 AC ABB08508;
 XX 23-MAY-2002 (first entry)
 DE Peptide related to revealing polypeptide from Escherichia coli.
 XX Carboxypeptidase Y propeptide.
 KW Unidentified.
 OS KR2000066750-A.
 XX 15-NOV-2000.
 XX 20-APR-1999; 99KR-0014065.
 XX 20-APR-1999; 99KR-0014065.
 XX (DONG-) DONG KOOK PHARM CO LTD.
 XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
 XX Chung BH, Oh GH, Hahn MS, Park JG;
 XX WPI; 2001-279683/29.
 XX

PT Method for revealing polypeptide from *Escherichia coli* with
 PT carboxypeptidase Y propeptide as the fusion partner, and recombinant
 PT vector and transformer thereof

XX Disclosure; Page 5; 18pp; Korean.

XX This invention relates to a method for revealing polypeptide from
 CC *Escherichia coli* with carboxypeptidase Y propeptide as the fusion
 CC partner. This peptide sequence is related to the invention.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 25

AAB99212
 ID AAB99212 standard; Peptide; 5 AA.

AC AAB99212;

DT 05-SEP-2001 (first entry)

DE Synthetic peptide.

XX SLC; cytostatic; cancer; chemokine; interleukin-2.

OS Synthetic.

PN WO200140311-A1.

XX 07-JUN-2001.

PF 27-NOV-2000; 2000WO-JP08325.

PR 30-NOV-1999; 99JP-0339315.

PA (SHIO) SHIONOGI & CO LTD.

PI Nakahara K, Sakata T;

DR WPI; 2001-374800/39.

XX Fusion protein containing interleukin-2 and chemokine SLC sequences and
 PT DNA encoding it for treatment of cancer

XX Disclosure; Page 66; 75pp; Japanese.

XX The present invention relates to fusion proteins containing chemokine SLC
 CC sequences fused at the N-termini of interleukin-2 sequences. The fusion
 CC proteins can be used for cancer therapy. The present sequence is a
 CC synthetic peptide which was used in the present invention.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 26

AAB97015

ID AAB97015 standard; peptide; 5 AA.

XX AAB97015;

DT 13-JUL-2001 (first entry)

XX Enterokinase recognition sequence #2.

DE Enterokinase recognition sequence; FLAG; epitope tag; cleavage site;
 KW identification polypeptide; protein purification;
 KW fusion protein production; antibody-based affinity chromatography;
 KW protein detection.

XX Unidentified.

PN WO200127293-A1.

XX 19-APR-2001.

PF 20-SEP-2000; 2000WO-US25693.

PR 08-OCT-1999; 99US-0415000.

PA (SIGMA-) SIGMA-ALDRICH CO.

PI Brizzard B, Hernan R;

DR WPI; 2001-282040/29.

XX New identification peptide containing many linked copies of an antigen
 PT domain, useful for purification and detection of recombinant fusion
 PT proteins

PS Disclosure; Page 10; Sipp; English.

XX The present sequence is recognised and cleaved by enterokinase. It is
 CC provided in a specification relating to identification polypeptides for
 CC purifying a target peptide. The identification polypeptides comprise
 CC multiple copies of an antigenic domain joined in tandem, a cleavable
 CC linking sequence and optionally a spacer domain. They are used for
 CC detection, production and purification of recombinant proteins, i.e.
 CC target peptide-identification polypeptide fusion proteins. The
 CC identification polypeptides make it possible to purify any protein, by
 CC antibody-based affinity chromatography, even where the protein is not
 CC antigenic and can not normally be purified by affinity methods. The
 CC presence of many epitopes in an identification polypeptide increases
 CC the detection limit of fusion proteins containing the identification
 CC polypeptide. The present sequence may be used as a cleavable site
 CC between the sequence of antigenic domains and the amino terminus of the
 CC target peptide. Enterokinase may then be used to liberate the
 CC target peptide from the identification polypeptide.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 27

AAB62307
 ID AAB62307 standard; peptide; 5 AA.

XX AAB62307;

DT 29-JUN-2001 (first entry)

DE Enterokinase cleavage site.

XX

KW Biosynthetic; antibody binding site; binding assay; drug targeting;
 KW biocatalysis; imaging; affinity purification; immunological; oncogenic;
 KW infectious disease; enterokinase; cytostatic; antimicrobial.
 XX Unidentified.
 OS
 XX
 PN US6207804-B1.
 XX
 PD 27-MAR-2001.
 XX
 XX 18-DEC-1995; 95US-0575724.
 XX
 PF 21-MAY-1987; 87US-0052800.
 XX
 PR 12-MAR-1992; 92US-0850228.
 PR 19-OCT-1993; 93US-0139901.
 PR 30-JUN-1998; 98US-0213671.
 XX
 XX (CURI-) CURIS INC.
 PA
 PI Huston JS, Oppermann H;
 XX
 XX WPI; 2001-280767/29.
 DR
 XX
 XX Novel biosynthetic antibody binding site having chimeric polypeptide
 PT which defines selective antigen binding region and has CDR and
 PT framework amino acid sequences homologous to different immunoglobulin
 PT molecules -
 PT
 PS Disclosure; Column 12; 26pp; English.
 XX
 XX The invention relates to a biosynthetic antibody binding site that
 CC comprises disulfide bonded heavy, light variable region (HVR, LVR)
 CC defining single, complete antigen (Ag1) binding site. HVR, LVR have
 CC polypeptide domain having sets of complementarity determining regions
 CC (CDR) and framework (FR) amino acid sequences homologous to first and
 CC second immunoglobulin, respectively. The biosynthetic antibody binding
 CC sites are useful in specific binding assays, affinity purification,
 CC biocatalysis, drug targeting, imaging immunological treatment of various
 CC oncogenic and infectious diseases. The present sequence represents a
 CC enterokinase cleavage site that can be used in the course of the
 CC invention.
 XX
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 22; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 DE 1 DDDDK 5
 DE Peptide #1.
 XX
 XX Coupling factor 6; CF6; antiinflammatory; antidiabetic; antiulcer;
 KW cardiant; hypotensive; antianginal; antiasthmatic; antirheumatic;
 KW antiarthritic; proton-translocating ATPase; prostaglandin;
 KW cytoplasmic PLA₂; cardiovascular disorder; diabetes;
 KW inflammatory disorder; rheumatoid arthritis; cerebrovascular disorder.
 XX
 XX Unidentified.
 OS
 XX WO200121205-A1.
 PN
 XX

PD 29-MAR-2001.
 XX
 PF 03-AUG-2000; 2000WO-JP05210.
 XX
 PR 17-SEP-1999; 99JP-0264687.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Osanai T, Magota K;
 XX
 DR WPI; 2001-266047/27.
 XX
 XX Coupling factor 6 useful in diagnosis of diseases relating to excess or
 PT lack of prostaglandin and cytoplasmic PLA₂ activity in blood, and in
 PT screening drugs for treating e.g. cardiovascular infarction -
 XX
 XX Example 1; Page 15; 70pp; Japanese.
 XX
 CC The present invention relates to human and rat coupling factor 6 (CF6)
 CC proteins (see AAB80963 and AAB80964). CF6 is a subunit of a
 CC proton-translocating ATPase found in mitochondria. CF6 can be used in
 CC diagnosis of diseases relating to excess or lack of prostaglandin and
 CC cytoplasmic PLA₂ activity, and in screening inhibitors and potentiators
 CC as drugs to treat e.g. cardiovascular infarction, hypertension,
 CC arteriosclerosis, angina pectoris, cardiac insufficiency, cerebrovascular
 CC disorders, hyperlipemia, diabetes, bronchiolitis, gastric ulcer, pregnant
 CC eclampsia, haemolytic uremia syndrome, thrombopenic purpura, inflammatory
 CC diseases like cerebral infarction, acute pancreatitis, asthma, ARDS
 CC (adult respiratory distress syndrome), and rheumatoid arthritis. The
 CC present sequence is a peptide used in the present invention.
 XX
 SQ Sequence 5 AA;
 Query Match 100.0%; Score 29; DB 22; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 1 DDDDK 5
 DE 1 DDDDK 5
 DE 31-MAY-2001 (first entry)
 XX
 XX Protease cleavage site inserted to remove His tag from GPVI sequence.
 DE
 XX Human; platelet membrane glycoprotein VI; GPVI; thrombolytic; therapy;
 KW vascular disease; thrombosis.
 KW
 XX Unidentified.
 OS
 XX WO200116321-A1.
 PN
 XX 08-MAR-2001.
 PD
 XX 01-SEP-2000; 2000WO-US23975.
 PF
 XX 01-SEP-1999; 99US-0152197.
 PR
 XX 08-OCT-1999; 99US-0158251.
 XX
 PA (SUKA) OTSUKA PHARM CO LTD.
 XX
 XX Tandon N, Sun B, Nakamura T, Yamamoto N;
 PI
 XX WPI; 2001-226691/23.
 DR
 XX Anti-thrombotic medicament, comprising a polypeptide having the

PT extracellular domain of platelet membrane glycoprotein VI or its
 PT variant, useful for treating a vascular disease and reducing platelet
 PT activation -

XX Disclosure; Page 36; 74pp; English.

XX The present sequence is a protease cleavage site, DDDDK. This site is
 CC inserted between the human platelet membrane glycoprotein VI (GPVI)
 CC sequence and the non-homologous portion of the fusion protein. This site
 CC is used to remove the histidine tag from the purified GPVI protein.
 CC The medicament comprising GPVI is useful for treating vascular disease,
 CC and for reducing platelet activation which involves contacting platelets
 CC with the medicament. The extracellular portion of GPVI is used
 CC therapeutically to attenuate platelet activation and aggregation and to
 CC treat thrombosis and other vascular diseases. Antibodies generated
 CC against GPVI are used as research and immunotherapeutic agents.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5

Db 1 DDDDK 5

RESULT 30

AAB35449
 ID AAB35449 standard; Peptide; 5 AA.

XX AC AAB35449;

XX DT 23-MAY-2001 (first entry)

XX DE Bovine enterokinase cleavage site.

XX KW Activatable neurotoxin; protease cleavage; H chain; L chain;
 KW clostridial neurotoxin.

XX OS Unidentified.

XX PN WO200114570-A1.

XX PD 01-MAR-2001.

XX PF 25-AUG-2000; 2000WO-US23427.

XX PR 25-AUG-1999; 99US-0150710.

XX PA (ALLR) ALLERGAN SALES INC.

XX PI Dolly JO, Li Y, Chan KC;

XX DR WPI; 2001-218454/22.

XX Novel isolated single-chain polypeptide derived from activatable
 PT recombinant clostridial neurotoxin useful as therapeutic agents,
 PT transporter molecules and adducts -

XX Claim 12; Page 9; 90pp; English.

XX The present invention describes an isolated single-chain protein
 CC comprising a therapeutic element, and a binding and translocation
 CC element. When exposed to a protease, the therapeutic element can be
 CC cleaved off. Examples of the binding and translocation element include
 CC the clostridial neurotoxin H and L chains. This is useful in the
 CC treatment of patients inoculated with the pentavalent BONT vaccine and
 CC for delivering the therapeutic benefits of neurotoxins to patients who
 CC are immunologically resistant to a given neurotoxin subtype, patients who
 CC may have a lower than average concentration of receptors to a given
 CC neurotoxin heavy chain binding group, or patients who may have a

CC protease-resistant variant of the membrane or vesicle toxin substrate.
 CC The present sequence was used to demonstrate the actions of the protein
 CC of the invention.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5

Db 1 DDDDK 5

RESULT 31

AAB48754
 ID AAB48754 standard; peptide; 5 AA.

XX AC AAB48754;

XX DT 09-MAR-2001 (first entry)

XX DE Fluorescent electrophoretic tag charged peptide linker #16.

XX KW Fluorescent electrophoretic tag; E-TAG; charged peptide linker;
 KW polymerase 5'-3' exonuclease activity; detector sequence;
 KW nucleic acid detection; single nucleotide polymorphism; SNP;
 KW allele; mutation.

XX OS Synthetic.

XX PN WO200066607-A1.

XX PD 09-NOV-2000.

XX PF 28-APR-2000; 2000WO-US11396.

XX PR 30-APR-1999; 99US-0303029.

XX PA (ACLA-) ACLARA BIOSCIENCES INC.

XX PI Singh S;

XX WPI; 2001-007201/01.

XX Detecting a DNA sequence, particularly a single nucleotide polymorphism
 PT using a pair of nucleotide sequences, a primer and an snp detection
 PT sequence having an electrophoretic tag -

XX Claim 22; Page 31; 76pp; English.

XX The invention relates to a method of detecting a DNA sequence,
 CC particularly one containing a single nucleotide polymorphism (SNP), in a
 CC sample. The method comprises amplifying a target DNA using a polymerase
 CC with 5'-3' exonuclease activity and a specific primer in the presence of
 CC an electrophoretically tagged detection sequence, which binds to the
 CC target DNA downstream of the primer binding site. The detection sequence
 CC contains an electrophoretic tag (E-TAG) attached to one of its
 CC nucleotides. The E-TAG comprises a fluorophore (or functional group for
 CC the attachment of a fluorophore) and a linker moiety such as a peptide
 CC (preferably charged), attached to a nucleotide of the detection
 CC sequence. The primer extension is performed, and the detector sequence
 CC bound to the target DNA is degraded by the 5'-3' exonuclease activity of
 CC the polymerase, with the release of the E-TAG which can then be
 CC separated into fractions and detected. When the method is used for the
 CC detection of SNPs, several detection sequences and upstream primers
 CC specific for each SNP (reagent pairs) are used, where each of the
 CC detection sequences has a different E-TAG attached to it. Having each
 CC SNP associated with its own E-TAG enables the SNPs present in the target
 CC DNA for which reagent pairs have been provided to be detected. The
 CC invention also provides for kits comprising several SNP detection
 CC sequences, and kits comprising several E-TAGS consisting of a fluorescer

CC and a moiety selected from a purine, pyrimidine, nucleoside, nucleotide
 CC of NTP, joined by a charged peptide linker (AAB48749-B48760). The method
 CC of the invention is useful for detecting at least one nucleic acid
 CC sequence or several SNPs in a target DNA sample, and provides for
 CC improved analysis of complex nucleic acid mixtures and for the
 CC simultaneous identification of entities such as sequences, SNPs, alleles,
 CC or mutations. The present sequence represents a charged peptide linker
 CC for use in an E-TAG of the invention.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 22; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 32

ABG71473
 ID ABG71473 standard; Peptide; 5 AA.

AC ABG71473;

XX 28-FEB-2003 (first entry)

XX Human parathyroxin (hPTH) related peptide #1.

XX Parathyroxin; hPTH.

XX Unidentified.

XX CN1353115-A.

XX 12-JUN-2002.

XX 13-NOV-2000; 2000CN-0133573.

XX 13-NOV-2000; 2000CN-0133573.

XX (MEIB/) MEI B.

XX Huang X, You Z, Li S;

XX WPI: 2002-751425/82.

XX N-PSDB; ABS7479.

XX Synthesis and expression of recombinant human parathyroxin gene -

XX Claim 4; Page 1 (Claims); 16pp; Chinese.

XX The invention relates to a DNA sequence containing a fragment of the
 CC human parathyroxin (hPTH) gene. The invention also relates to splicing
 CC artificially synthesised oligodeoxynucleotide chains by PCR and preparing
 CC fusion or non-fusion expressed engineering bacteria of recombination
 CC protein hPTH through integrating the gene in different expression
 CC carriers. This sequence represents a human parathyroxin gene related
 CC peptide.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 33

ABB98260

ID ABB98260 standard; protein; 5 AA.

XX ABB98260;

XX 06-NOV-2002 (first entry)

XX Enteropeptidase enterokinase peptide binding sequence.

XX Trypsinogen; enzyme; enteropeptidase; enterokinase.

XX Synthetic.

XX WO200261064-A2.

XX 08-AUG-2002.

XX 01-FEB-2002; 2002WO-EP01072.

XX 01-FEB-2001; 2001EP-0102342.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Mueller R, Glaser S, Gelpel F, Thalhofer J, Rexer B, Schneider C;

XX Ratka M, Ronning S, Eckstein H, Giessel C;

XX WPI; 2002-619246/66.

XX Preparation of recombinant trypsin, useful e.g. in peptide sequencing,
 PT by expressing truncated trypsinogen under conditions that prevent
 PT autocatalytic cleavage -

XX Claim 15; Page 33; 45pp; German.

XX The invention relates to preparation of recombinant trypsin (I) by
 CC culturing host cells that have been transformed with recombinant nucleic
 CC acid (II) that expresses secretable trypsinogen (Ia) containing an
 CC enterokinase recognition site in the propeptide sequence (PPS). Culture
 CC conditions prevent autocatalytic cleavage of PPS. The expression product
 CC is isolated from the medium and PPS cleaved to form active (I), from
 CC which uncleaved is optionally separated. (I) is used to cleave peptides
 CC to short fragments for sequencing; to release adherent cells from culture
 CC surfaces; to cleave fusion proteins; for activation of propeptides, e.g.
 CC conversion or recombinant proinsulin to insulin and as a component of
 CC pharmaceuticals. The present sequence is that of the enteropeptidase
 CC enterokinase peptide binding sequence useful in generating the synthetic
 CC trypsinogen of the invention.

XX Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 34

ABJ05330

ID ABJ05330 standard; Peptide; 5 AA.

XX ABJ05330;

XX 08-NOV-2002 (first entry)

XX Human enterokinase peptide cleavage sequence.

XX Target peptide production; fusion peptide; protease-susceptible linker;
 KW parathyroid hormone; PTH; high expression rate;
 KW pharmaceutical application.

XX OS Homo sapiens.
XX PN WO200236762-A1.
XX PD 10-MAY-2002.
XX PF 29-OCT-2001; 2001WO-JP09476.
XX PR 30-OCT-2000; 2000JP-0331170.
XX PR 27-JUN-2001; 2001JP-0195522.
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX PI Yamada T, Suenaga M;
XX WPI; 2002-417275/44.
XX DR N-PSDB; ABT06800.
XX PT Production of target peptide comprises cleavage of fusion peptide with
PT parathyroid hormone peptide for efficient manufacture of target peptide
PT without the need to remove N-terminal methionine .
XX PS Claim 4; Page 14; 103pp; Japanese.
XX CC The invention comprises a method of producing a target peptide. The
CC C-terminal end of the target peptide is fused via a protease-susceptible
CC linker to parathyroid hormone (PTH) residues 1-34. The method of the
CC invention is useful for the clean and efficient production of a target
CC peptide at a high expression rate on an industrial scale without the need
CC to remove the N-terminal methionine from the product. The peptides
CC produced by the method of the invention are suitable for pharmaceutical
CC and other uses. The present peptide sequence was used in the invention.
XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 35
AAE25730
ID AAE25730 standard; peptide; 5 AA.
XX AC AAE25730;
XX DT 04-NOV-2002 (first entry)
XX DE Human and bovine trypsin activation peptide.
XX KW Chymotrypsin-like polypeptide; elastase-like polypeptide; amidolytic;
KW Soli E2; trypsin; enzyme; human; bovine.
XX OS Homo sapiens.
XX OS Bos sp.
XX PN US6399759-B1.
XX PD 04-JUN-2002.
XX PF 25-MAY-2000; 2000US-0578303.
XX PR 27-MAY-1999; 99US-136331P.
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX PI Travis J, Whitworth ST, Blum MS;
XX

DR WPI; 2002-546298/58.
XX Novel isolated nucleic acid fragment encoding a serine protease, useful
PT in identifying molecules that inhibit its activity, and in the
PT development of insecticides to control fire ant infestation .
XX Example 2; Column 25; 31pp; English.
XX The invention relates to polypeptides including chymotrypsin-like
CC polypeptides and elastase-like polypeptides, having amidolytic activity
CC for cleavage of a peptide bond present in a target polypeptide. Nucleic
CC acid encoding the polypeptides of the invention is useful to transform
CC a host cell which optionally includes one or more marker sequences,
CC which encodes a polypeptide that inactivates or detects a compound in
CC the growth medium. The encoded polypeptide can be used to develop an
CC inhibitor that reduces the amidolytic activity of the polypeptide.
CC and the inhibitor can then be used to reduce the amount of solid food
CC digested by a Solenopsis invicta 4th instar larva or colony. The
CC present sequence is human and bovine trypsin activation peptide used
CC in the exemplification of the invention.
XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 36
AAE23661
ID AAE23661 standard; peptide; 5 AA.
XX AC AAE23661;
XX DT 10-SEP-2002 (first entry)
XX DE Enterokinase cleavage site.
XX KW Kringle peptide; angiogenesis-associated disease; psoriasis; cancer;
KW arthritis; macular degeneration; endothelial cell proliferation;
KW diabetic retinopathy; cytostatic; enterokinase cleavage site.
XX OS Unidentified.
XX PN WO200226782-A2.
XX PD 04-APR-2002.
XX PF 27-SEP-2001; 2001WO-US42423.
XX PR 29-SEP-2000; 2000US-0675226.
XX PR 31-AUG-2001; 2001US-0942704.
XX PA (ABBO) ABBOTT LAB.
XX PI Henkin J, Davidson DJ;
XX WPI; 2002-452273/48.
XX Using kringle peptides conjugated to functionalized polymers, e.g.
PT ethoxypolyethylene, polyethylene glycol or methoxypolyethylene glycol,
PT to treat angiogenic disorders e.g. cancer, macular degeneration and
PT arthritis .
XX Example 1; Page 16; 34pp; English.
XX The present invention relates to conjugated kringle peptide fragments
CC consisting of a functionalised kringle peptide fragment chemically
CC coupled to a functionalised polymer. The conjugated kringle peptides

CC may be administered to a patient for treating angiogenesis-associated
CC diseases such as psoriasis, cancer, arthritis, macular degeneration,
CC diabetic retinopathy and for inhibiting endothelial cell proliferation.
CC The present sequence is enterokinase cleavage site. This sequence is
CC used in the exemplification of the invention.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 37

AAU98272
ID AAU98272 standard; peptide; 5 AA.

XX AC AAU98272;

XX DT 15-AUG-2002 (first entry)

XX DE Synthetic adaptor peptide.

XX KW Peptide production; fusion peptide; histidine tag; enterokinase; ACE.

XX OS Synthetic.

XX PN JP2002119284-A.

XX PD 23-APR-2002.

XX PF 13-OCT-2000; 2000JP-0313609.

XX PR 13-OCT-2000; 2000JP-0313609.

XX PA (TOHA-) TOYO HAKKO KK.

XX DR WPI; 2002-475328/51.

XX PT A process for preparation of short length peptides by culture of a
transformant with an expression vector, followed by cleavage with an
enzyme recognising the selective enzyme cleavage site -

XX PS Example 1; Page 9; 12pp; Japanese.

XX CC The invention relates to a process for preparation of short length
peptides by culture of a transformant with an expression vector,
prepared by insertion of a gene encoding the fused peptide with the
short peptide and the selective enzyme cleavage site, containing 2 or
more genes encoding for a fused peptide with a short length peptide,
particularly containing a histidine tag, and a selective enzyme
cleavage site, particularly the site with cleaved enterokinase,
followed by cleavage with the enzyme recognising the selective enzyme
cleavage site. The method involves selective cleavage of long chain
peptides to give the aimed short length peptide. The present
sequence is a synthetic peptide which may be fused with a peptide derived
from the ACE protein (not defined) and used to demonstrate the method
of the invention.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 38

AAU97026

ID AAU97026 standard; Peptide; 5 AA.

XX AC AAU97026;

XX DT 13-AUG-2002 (first entry)

XX DE Cattle enteropeptidase catalysing subunit-associated peptide #3.

XX KW Cattle enteropeptidase catalysing subunit; intestinal tissue;

XX KX serine protease; enzyme.

XX OS Unidentified.

XX PN CN1332248-A.

XX PD 23-JAN-2002.

XX PF 10-JUL-2001; 2001CN-0113770.

XX PR 10-JUL-2001; 2001CN-0113770.

XX PA (UYNA-) UNIV NANJING.

XX PI Hua Z, Yuan L;

XX DR WPI; 2002-292900/34.

XX PT Cattle enteropeptidase catalysing subunit gene and its gene engineering
production process -

XX PS Disclosure; Page 1 (Disclosure); 9pp; Chinese.

XX CC The invention relates to a cattle enteropeptidase catalysing subunit.
The gene may be obtained from cattle's intestinal tissue by gene cloning
technology. The cattle enteropeptidase catalysing subunit may be prepared
through gene engineering process and may be identified specifically as
a serine protease. The enteropeptidase may be used as a protease for
the specific cleavage of recombinant fusion proteins. The present
sequence represents the amino acid sequence of an cattle enteropeptidase
catalysing subunit-associated peptide.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 39

AAE20192

ID AAE20192 standard; peptide; 5 AA.

XX AC AAE20192;

XX DT 18-JUN-2002 (first entry)

XX DE Enterokinase cleavage peptide.

XX KW Iron-regulated protein; vaccine; bactericide; infection;
transferrin binding protein; TBP; enterokinase.

XX OS Unidentified.

XX PN US2002025318-A1.

XX PD 28-FEB-2002.

XX PF 14-JUN-1999; 99US-0332226.
XX XX
PR 05-NOV-1992; 92US-0973336.
PR 20-SEP-1993; 93US-0124254.
PR 23-DEC-1994; 94US-0363124.
PR 23-AUG-1990; 90US-0572187.
XX XX
PA (UYNC-) UNIV NORTH CAROLINA.
XX PI Sparling PF, Nau Cornelissen C;
XX XX WPI; 2002-267521/31.
DR WPI; 2002-267521/31.
XX XX
PT Iron-regulated proteins found in *Neisseria gonorrhoeae* or *Neisseria*
PT meningitidis outer membranes, useful for diagnosing and vaccinating
PT against *N. gonorrhoeae* and *N. meningitidis*.
XX XX
PS Disclosure; Page 5; 39pp; English.
XX XX
CC The present invention relates to iron-regulated transferrin binding
CC proteins (and functional analogues) found in *Neisseria gonorrhoeae* or
CC *Neisseria meningitidis* outer membranes. The proteins may be used for
CC detection or vaccination against infections caused by *N. meningitidis*
CC or *N. gonorrhoeae*. The antibodies are useful for inhibiting the growth
CC of *N. meningitidis* and/or *N. gonorrhoeae* and are also useful in
CC controlling infections of these pathogens. The antibodies are also
CC useful for detecting the presence of *N. gonorrhoeae* or *N. meningitidis*
CC in samples. The present invention is enterokinase cleavage peptide
CC used in the invention.
XX XX
SQ Sequence 5 AA;
Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DDDDK 5
Db 1 DDDDK 5
IIII
IIII
RESULT 40
AAU78925
ID AAU78925 standard; Peptide; 5 AA.
XX AC AAU78925;
XX XX
DT 18-JUN-2002 (first entry)
XX XX
DE Enterokinase specific recognition sequence cleavable peptide linker.
XX XX
KW Nuclear localisation signal; NLS; protein delivery; enterokinase;
KW fusion protein; membrane penetrating peptide.
XX XX
OS Unidentified.
XX XX
PN WO200218572-A2.
XX XX
PD 07-MAR-2002.
XX XX
PF 23-AUG-2001; 2001WO-US26421.
XX XX
PR 25-AUG-2000; 2000US-227647P.
PR 07-FEB-2001; 2001GB-0003110.
XX XX
PA (AVET) AVENTIS PHARM INC.
XX XX
PI Guo Y, Morse CC, Yao Z, Keesler GA;
XX XX WPI; 2002-304256/34.
DR WPI; 2002-304256/34.
XX XX
PT New fusion proteins comprising membrane penetrating peptides, useful as

PT in vivo, ex vivo or in vitro intracellular carriers or delivery devices
PT for a compound of interest (e.g. peptide, protein, chemical entity,
PT nucleic acid).
XX XX
PS Disclosure; Page 21; 45pp; English.
XX XX
CC This invention relates to a novel fusion protein, which comprises a
CC membrane penetrating peptide attached to a compound of interest.
CC The membrane penetrating peptide of the fusion protein is derived from a
CC nuclear localisation signal and may be the nuclear localisation signal
CC from human period protein hPER1. The fusion protein is useful for
CC delivery of a compound of interest into a cell. The fusion protein is
CC useful as in vivo, ex vivo or in vitro intracellular delivery devices
CC for a compound of interest (e.g. peptide, protein, chemical entity,
CC nucleic acid). In particular, the polypeptides are useful as protein
CC carriers for delivery of compounds to cells. The present sequence
CC represents a cleavable peptide linker sequence recognised by
CC enterokinase. This sequence may be used to create a fusion protein of
CC the invention.
XX XX
SQ Sequence 5 AA;
Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DDDDK 5
Db 1 DDDDK 5
IIII
IIII
RESULT 41
ABB06267
ID ABB06267 standard; Peptide; 5 AA.
XX AC ABB06267;
XX XX
DT 23-MAY-2002 (first entry)
XX XX
DE Peptide sequence.
XX XX
KW Rice; phospholipid hydroperoxide glutathione peroxidase; PHGPx;
KW glutathione phosphatide hydroperoxidase; senility; transgenic plant;
KW plant; enzyme.
XX XX
OS Synthetic.
XX XX
PN CN1324817-A.
XX XX
PD 05-DEC-2001.
XX XX
PF 19-MAY-2000; 2000CN-0109313.
XX XX
PR 19-MAY-2000; 2000CN-0109313.
XX XX
PA (UYQI) UNIV QINGHUA.
XX XX
PI Liu J, Li W, Zhao N;
XX XX WPI; 2002-217502/28.
DR WPI; 2002-217502/28.
XX XX
PT Rice glutathione phospholipid hydrogen peroxidase gene, protein and
PT their application.
XX XX
PS Example 5; Page 14 (Disclosure); 24pp; Chinese.
XX XX
CC The present invention describes rice phospholipid hydroperoxide
CC glutathione peroxidase (PHGPx), also called glutathione phosphatide
CC hydroperoxidase). The invention also provides a recombinant expression
CC vector containing the PHGPx gene, and a host cell containing the
CC expression vector. The rice PHGPx gene can be used for protecting cell
CC membrane injury due to phosphatide hydroperoxidation. It can also be
CC used for making transgenic plants delaying senility, specially making

CC transgenic rice capable of delaying senility and raising photosynthetic
CC efficiency. The present sequence represents a peptide which is used an
CC example from the present invention.
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 1 DDDDK 5

RESULT 42
AAB47905
ID AAB47905 standard; peptide; 5 AA.
XX
XX AAB47905;
XX AC
XX DT
XX 16-MAY-2002 (first entry)
XX Linker peptide enterokinase recognition sequence.
XX
XX Recursive ensemble mutagenesis; REM; cell viability; optical signal;
KW high-throughput screening; antimicrobial compound; antibiotic.
XX
XX Synthetic.
XX WO200206517-A2.
XX
XX 24-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23004.
XX
XX 19-JUL-2000; 2000US-219179P.
XX
XX (KAIR-) KAIROS SCI INC.
XX
XX Bylina EJ, Coleman WJ, Youvan DC;
XX WPI; 2002-179801/23.
XX
XX Screening compounds affecting cell viability e.g. for identifying
PT antimicrobial compounds, comprises determining if induced transformed
PT cell colonies have a desired signal when contacted with a viability
PT indicator -
XX
XX Disclosure; Page 22; 56pp; English.
XX
XX The sequences given in AAB47904-05 represent fragments of a linker
CC peptide which was expressed in a recursive ensemble mutagenesis (REM)
CC cassette of the invention. The REM cassette contained an antimicrobial
CC peptide library cloned into a plasmid cassette contained the chimeric
CC Lpp-OmpA display protein. The cassette may be used in the method of the
CC invention for determining whether a compound affects cell viability by:
CC (a) exposing colonies of cells (CC) on a support surface to inducing
CC conditions, where the cells have been transformed with an expression
CC library encoding candidate compounds;
CC (b) contacting CC with a viability indicator that produces an optical
CC signal indicative of cell viability; and
CC (c) determining if a colony has a desired optical signal.
CC The method is useful for determining whether a compound affects cell
CC viability. It is useful for high-throughput screening to identify
CC antimicrobial compounds and in drug discovery. The antimicrobial
CC compounds are useful in the pharmaceutical industry, and provide an
CC additional new class of antibiotic compounds to fight infectious
CC diseases. The method is useful for assaying the authentic peptide
CC sequences contained in an expression library for antimicrobial
CC activity, for distinguishing dead cells (expressing active sequences)
CC from living cells (expressing inactive or less active sequences), to
CC identify novel antimicrobial peptide sequences, including highly potent

CC molecules, resulting in a large number of new antimicrobial lead
CC compounds that are active against a broad range of bacteria or other
CC microorganisms, and for screening all types of antibiotic compounds,
CC including libraries of low molecular weight molecules produced by
CC metabolic engineering and artificial synthesized libraries in
CC solid-phase arrays.
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 23; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 1 DDDDK 5

RESULT 43
ABP72097
ID ABP72097 standard; peptide; 5 AA.
XX
XX ABP72097;
XX AC
XX 21-MAY-2003 (first entry)
XX
XX Enterokinase target sequence.
XX
XX Enterokinase; extracellular target molecule; gene fusion protein.
XX
XX Synthetic.
XX WO2003008436-A2.
XX
XX 30-JAN-2003.
XX
XX 18-JUL-2002; 2002WO-US23375.
XX
XX 19-JUL-2001; 2001US-306924P.
XX
XX (ICOG-) ICOGEN CORP.
XX
XX Hagen FS, Woodbury RG, Oort PJ;
XX WPI; 2003-239306/23.
XX
XX Identifying peptides that specifically bind to extracellular target
PT molecules comprises expressing peptide libraries on the surface of
PT mammalian host cells and peptide screening under substantially
PT physiological conditions -
XX
XX Disclosure; Page 6; 69pp; English.
XX
XX The invention relates to a novel method for identifying peptides that
CC bind to extracellular target molecules, comprising introducing an
CC expression library comprising a plurality of oligonucleotides into
CC mammalian host cells, which express and display the peptide on an
CC extracellular cell surface, and contacting the host cells with an
CC extracellular target molecule under physiological conditions or in a
CC substantially undiluted biological fluid. The methods of the invention
CC are useful for identifying peptides that specifically bind to
CC extracellular target molecules. The vectors are useful for expressing
CC gene fusion proteins and for targeting the fusion proteins to the
CC extracellular cell surface. The present sequence represents a peptide of
CC the invention.
CC Note: The sequence described on page 22 as SEQ ID NO 3 varies from the
CC sequence given in the sequence listing, and is described as SEQ ID NO 4
CC on page 24. The sequence described as SEQ ID NO 3 on page 24 is SEQ ID
CC NO 2, and the sequence described as SEQ ID NO 5 is SEQ ID NO 4.
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 24; Length 5;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||

Db 1 DDDDK 5

RESULT 44

ABG76088

ID ABG76088 standard; peptide; 5 AA.

AC ABG76088;

DT 16-MAY-2003 (first entry)

DE Enterokinase cleavage site.

XX Enterokinase cleavage site; angiogenesis; primary solid tumour; cancer;
KW metastatic solid tumour; carcinoma; sarcoma; lymphoma; ocular disease;
KW autoimmune disease prophylaxis; rheumatoid arthritis; immune arthritis;
KW diabetic retinopathy; psoriasis; macular degeneration; skin disease;
KW abnormal eye neovascularisation; blood vessel disease; haemangioma;
KW Osler's disease; angiofibroma; plaque neovascularisation; telangiectasia;
KW capillary proliferation within atherosclerotic plaque; wound granulation;
KW myocardial angiogenesis; haemophilic joint; atherosclerosis; ulcer;
KW cat scratch disease; endothelial cell stimulation; Crohn's disease;
KW placenta establishment; ovulation inhibition; birth control;
KW endothelial cell proliferation.

XX Unidentified.

XX OS

XX PN US2002159992-A1.

XX PD 31-OCT-2002.

XX PF 28-SEP-2001; 2001US-0967386.

XX PR 29-SEP-2000; 2000US-236550P.

XX PA (HENK/) HENKIN J.

XX PA (DAVI/) DAVIDSON D J.

XX PI Henkin J, Davidson DJ;

XX DR WPI; 2003-298673/29.

XX Conjugated kringle peptide fragment of plasminogen, useful for treating
PT angogenic disease, e.g. cancer, comprises functionalized kringle
PT peptide fragment chemically coupled to functionalized polymer -

XX Example 1; Page 7; 17pp; English.

XX The invention relates to a conjugated kringle peptide fragment which
CC comprises a functionalised kringle peptide fragment chemically coupled to
CC a functionalised polymer. The fragment is used for treating a patient in
CC need of anti-angiogenic therapy, comprising administering a conjugated
CC kringle peptide, preferably kringles 4-5 of plasminogen. The disease can
CC be primary and metastatic solid tumours; carcinomas; sarcomas; lymphomas;
CC prophylaxis of autoimmune disease e.g. rheumatoid arthritis and immune
CC arthritis; ocular disease e.g. diabetic retinopathy and macular
CC degeneration; abnormal neovascularisation conditions of the eye; skin
CC diseases e.g. psoriasis; blood vessel disease e.g. haemangiomas and
CC capillary proliferation within atherosclerotic plaques; Osler's disease;
CC myocardial angiogenesis; plaque neovascularisation; telangiectasia;
CC haemophilic joints; angiofibroma; wound granulation; excessive or
CC abnormal stimulation of endothelial cells e.g. Crohn's disease;
CC atherosclerosis; cat scratch disease; ulcers. The fragment can also be
CC used as a birth control agent which inhibits ovulation and establishment
CC of the placenta. The fragment is also used for inhibiting endothelial
CC cell proliferation in an individual and in vitro. The invention has an
CC improved pharmacokinetic activity and is easily and cost-effectively
CC made. The present sequence represents the amino acid sequence of an

CC enterokinase cleavage site used in the generation of kringle fusion
XX proteins.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 24; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

 |||||

Db 1 DDDDK 5

RESULT 45

ABU60856

ID ABU60856 standard; Peptide; 5 AA.

XX AC

XX ABU60856;

XX DT 06-MAY-2003 (first entry)

XX DE Peptide production by gene recombination associated peptide #40.

XX KW Peptide production; low-molecular peptide; KiSS-1; GPR8 ligand;

XX gene recombination.

XX OS Synthetic.

XX PN WO200292829-A1.

XX PD 21-NOV-2002.

XX PF 16-MAY-2002; 2002WO-JP04735.

XX PR 17-MAY-2001; 2001JP-0147341.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Nishimura O, Suenaga M, Ito T, Kitada C;

XX DR WPI; 2003-129302/12.

XX Process for producing peptides e.g. KiSS-1 peptide and GPR8 ligand for
PT subsequent applications by gene recombination technique through tandem
PT repeats to provide precursor protein with specific cleavage sites -

XX Claim 8; Page 41; 87pp; Japanese.

XX The invention describes a method of producing a peptide comprising the
CC excision of the N and C-terminals of a target peptide with enzymes or
CC chemically through the attached cleavage sites repeated by ligation in a
CC precursor protein. The method is for producing (low-molecular) peptides
CC e.g. KiSS-1 peptide and GPR8 ligand for subsequent applications by the
CC gene recombination technique through tandem repeats to provide
CC a precursor protein with specific cleavage sites. With this method,
CC peptide production can be carried out easily to provide large quantities
CC of the required peptides. This is the amino acid sequence of a peptide
CC associated with the peptide production method of the invention.

SQ Sequence 5 AA;

Query Match 100.0%; Score 29; DB 24; Length 5;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

 |||||

Db 1 DDDDK 5

Search completed: October 14, 2003, 13:14:21

Job time : 35.6364 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 13.1818 Seconds
(without alignments)
16.049 Million cell updates/sec

Title: US-09-856-050-19_COPY_36_40

Perfect score: 29

Sequence: 1 DDDDK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	5	1 US-07-816-679A-7	Sequence 7, Appl
2	29	100.0	5	1 US-08-197-496A-21	Sequence 21, Appl
3	29	100.0	5	1 US-08-131-365B-49	Sequence 49, Appl
4	29	100.0	5	1 US-08-275-370-26	Sequence 26, Appl
5	29	100.0	5	1 US-08-367-968-26	Sequence 26, Appl
6	29	100.0	5	1 US-08-294-434-2	Sequence 2, Appl
7	29	100.0	5	1 US-08-457-166-2	Sequence 2, Appl
8	29	100.0	5	1 US-08-200-900A-34	Sequence 34, Appl
9	29	100.0	5	2 US-08-665-484-26	Sequence 26, Appl
10	29	100.0	5	2 US-07-963-538B-11	Sequence 11, Appl
11	29	100.0	5	2 US-08-668-123-49	Sequence 49, Appl
12	29	100.0	5	2 US-08-595-043A-13	Sequence 13, Appl
13	29	100.0	5	2 US-09-016-366A-29	Sequence 29, Appl
14	29	100.0	5	2 US-08-595-868C-16	Sequence 16, Appl
15	29	100.0	5	3 US-08-481-435-41	Sequence 41, Appl
16	29	100.0	5	3 US-08-483-749A-33	Sequence 33, Appl
17	29	100.0	5	3 US-08-997-918-56	Sequence 56, Appl
18	29	100.0	5	3 US-08-997-532B-11	Sequence 11, Appl
19	29	100.0	5	3 US-08-888-381-5	Sequence 5, Appl
20	29	100.0	5	3 US-08-927-128-6	Sequence 6, Appl
21	29	100.0	5	3 US-09-020-880-37	Sequence 37, Appl
22	29	100.0	5	3 US-09-197-801-15	Sequence 15, Appl
23	29	100.0	5	3 US-09-551-028-15	Sequence 15, Appl
24	29	100.0	5	3 US-08-938-595-3	Sequence 3, Appl
25	29	100.0	5	3 US-08-727-153-3	Sequence 3, Appl
26	29	100.0	5	3 US-09-139-819A-16	Sequence 16, Appl
27	29	100.0	5	3 US-08-840-466A-26	Sequence 26, Appl

Sequence 14, Appl
Sequence 3, Appl
Sequence 37, Appl
Sequence 29, Appl
Sequence 26, Appl
Sequence 16, Appl
Sequence 14, Appl
Sequence 143, Appl
Sequence 14, Appl
Sequence 14, Appl
Sequence 39, Appl
Sequence 6, Appl
Sequence 9, Appl
Sequence 7, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 21, Appl
Sequence 35, Appl

28 29 100.0 5 4 US-09-260-038B-14
29 29 100.0 5 4 US-09-415-000-3
30 29 100.0 5 4 US-09-101-544-37
31 29 100.0 5 4 US-09-578-303-29
32 29 100.0 5 4 US-09-696-188B-26
33 29 100.0 5 4 US-09-750-913-16
34 29 100.0 5 4 US-09-635-923-14
35 29 100.0 5 4 US-07-757-022B-143
36 29 100.0 5 4 US-09-180-422B-14
37 29 100.0 5 4 US-09-487-716A-14
38 29 100.0 5 4 US-09-059-625-39
39 29 100.0 5 4 US-09-464-152A-6
40 29 100.0 5 4 US-09-185-852-9
41 29 100.0 5 5 PCT-US92-11270-7
42 29 100.0 5 5 PCT-US93-06591-2
43 29 100.0 5 5 PCT-US95-15800-2
44 29 100.0 6 1 US-08-243-082-21
45 29 100.0 6 1 US-08-200-900A-35

ALIGNMENTS

RESULT 1
US-07-816-679A-7
; Sequence 7, Application US/07816679A
; Patent No. 5298599
; GENERAL INFORMATION:
; APPLICANT: Rezale, Alireza
; APPLICANT: Esmon, Charles T.
; APPLICANT: Morrissey, James H.
; TITLE OF INVENTION: Expression and Purification of
; TITLE OF INVENTION: Recombinant Soluble Tissue Factor
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street, Suite 3200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/816,679A
; FILING DATE: 19920103
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/730040
; FILING DATE: 12-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/292447
; FILING DATE: 30-DEC-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/683682
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 1..5
; OTHER INFORMATION: /note= "Enterokinase Cleavage Site"
US-07-816-679A-7

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 2
US-08-197-496A-21
; Sequence 21, Application US/08197496A
; Patent No. 5480797
; GENERAL INFORMATION:
; APPLICANT: FRANCHILLA, ANTONIO T.
; APPLICANT: HAGIYA, MICHIO
; APPLICANT: STARZL, THOMAS E.
; TITLE OF INVENTION: AN AUGMENTER OF LIVER REGENERATION (ALR)
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/197,496A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: 6137/203420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-197-496A-21

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 3
US-08-131-365B-49
; Sequence 49, Application US/08131365B
; Patent No. 5527690

;
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-Oct-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-131-365B-49

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 4
US-08-275-370-26
; Sequence 26, Application US/08275370
; Patent No. 5550037
; GENERAL INFORMATION:
; APPLICANT: FRANCHILLA, ANTONIO T.
; APPLICANT: HAGIYA, MICHIO
; APPLICANT: STARZL, THOMAS E.
; TITLE OF INVENTION: AN AUGMENTER OF LIVER REGENERATION (ALR): HUMAN
; TITLE OF INVENTION: AND RAT
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/275,370
;/ FILING DATE:
;/ CLASSIFICATION: 530
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: SCOTT, WATSON T.
;/ REGISTRATION NUMBER: 26,581
;/ REFERENCE/DOCKET NUMBER: 1140/204509
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 202-861-3000
;/ TELEFAX: 202-822-0944
;/ TELEX: 6714627 CUSH
;/ INFORMATION FOR SEQ ID NO: 26:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ US-08-275-370-26

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 1 DDDDK 5

RESULT 5
US-08-367-968-26
; Sequence 26, Application US/08367968
; Patent No. 5607844
; GENERAL INFORMATION:

;/ APPLICANT: FRANCAVILLA, ANTONIO T.
;/ APPLICANT: HAGIYA, MICHIO
;/ APPLICANT: STARZEL, THOMAS E.
;/ TITLE OF INVENTION: AN AUGMENTER OF LIVER REGENERATION (ALR): HUMAN
;/ TITLE OF INVENTION: AND RAT
;/ NUMBER OF SEQUENCES: 33
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: CUSHMAN DARBY & CUSHMAN
;/ STREET: 1100 NEW YORK AVENUE, N.W.
;/ CITY: WASHINGTON
;/ STATE: DC
;/ COUNTRY: USA
;/ ZIP: 20005-3918

;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/367,968
;/ FILING DATE:
;/ CLASSIFICATION: 435

;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: KOKULIS, PAUL N.
;/ REGISTRATION NUMBER: 16,773
;/ REFERENCE/DOCKET NUMBER: 1140/215562
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 202-861-3000
;/ TELEFAX: 202-822-0944
;/ TELEX: 6714627 CUSH

;/ INFORMATION FOR SEQ ID NO: 26:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ US-08-367-968-26

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 1 DDDDK 5

RESULT 6

US-08-294-434-2
; Sequence 2, Application US/08294434
; Patent No. 5635371
; GENERAL INFORMATION:
;/ APPLICANT: Stout, Jay
;/ APPLICANT: Wagner, Fred W.
;/ APPLICANT: Coolidge, Thomas R.
;/ APPLICANT: Holmquist, Barton
;/ TITLE OF INVENTION: CHEMICAL METHOD FOR SELECTIVE
;/ TITLE OF INVENTION: MODIFICATION OF THE N- AND/OR C-TERMINAL AMINO ACID
;/ TITLE OF INVENTION: ALPHA-CARBON REACTIVE GROUP OF A RECOMBINANT
;/ TITLE OF INVENTION: POLYPEPTIDE OR A PORTION THEREOF
;/ NUMBER OF SEQUENCES: 15
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Merchant & Gould
;/ STREET: 3100 No. 5635371west Center
;/ CITY: Minneapolis
;/ STATE: MN
;/ COUNTRY: USA
;/ ZIP: 55402

;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/294,434
;/ FILING DATE:
;/ CLASSIFICATION: 435

;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/912,798
;/ FILING DATE: 13-JUL-1992
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Nelson, Albin J.
;/ REGISTRATION NUMBER: 28,650
;/ REFERENCE/DOCKET NUMBER: 8648.29-US01
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 612-332-5300
;/ TELEFAX: 612-332-9081

;/ INFORMATION FOR SEQ ID NO: 2:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-294-434-2

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 1 DDDDK 5

RESULT 7

US-08-457-166-2
; Sequence 2, Application US/08457166
; Patent No. 5656456
; GENERAL INFORMATION:

```
; APPLICANT: Stout, Jay
; APPLICANT: Wagner, Fred W.
; APPLICANT: Coolidge, Thomas R.
; APPLICANT: Holmquist, Barton
; TITLE OF INVENTION: CHEMICAL METHOD FOR SELECTIVE
; TITLE OF INVENTION: MODIFICATION OF THE N- AND/OR C-TERMINAL AMINO ACID
; TITLE OF INVENTION: ALPHA-CARBON REACTIVE GROUP OF A RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDE OR A PORTION THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5656456west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,166
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/091,751
; FILING DATE:
; APPLICATION NUMBER: US 07/912,798
; FILING DATE: 13-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, Albin J.
; REGISTRATION NUMBER: 28,650
; REFERENCE/DOCKET NUMBER: 8648.35-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-457-166-2

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 8
US-08-200-900A-34
; Sequence 34, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meindert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-200-900A-34

Query Match 100.0%; Score 29; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 9
US-08-665-484-26
; Sequence 26, Application US/08665484
; Patent No. 5811397
; Patent No. 5811397 5780430
; GENERAL INFORMATION:
; APPLICANT: FRANCAVILLA, ANTONIO T.
; APPLICANT: HAGIYA, MICHIO
; APPLICANT: STARZL, THOMAS E.
; TITLE OF INVENTION: AN AUGMENTER OF LIVER REGENERATION (ALR):
; TITLE OF INVENTION: HUMAN AND RAT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,484
; FILING DATE: 12-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,968
; FILING DATE: 03-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 1140/215562
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
```

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-665-484-26

Query Match 100.0%; Score 29; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 10
US-07-963-538B-11
; Sequence 11, Application US/07963538B
; Patent No. 5851983
; GENERAL INFORMATION:
; APPLICANT: SUGIYAMA, TAKASHI
; APPLICANT: KAMIMURA, TAKASHI
; APPLICANT: MASUDA, KENICHI
; APPLICANT: OKADA, MASAHIRO
; APPLICANT: OHTSUKA, EIKO
; APPLICANT: IMAIZUMI, ATSUSHI
; APPLICANT: WATANABE, KUNIHITO
; APPLICANT: SUGA, TETSUYA
; APPLICANT: MATSUMOTO, YOHICHI
; APPLICANT: TAKEUCHI, AKIKO
; TITLE OF INVENTION: ELASTASE INHIBITORY POLYPEPTIDE AND
; TITLE OF INVENTION: PROCESS FOR PRODUCTION THEREOF BY RECOMBINANT GENE
; TITLE OF INVENTION: TECHNOLOGY
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD LLP
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/963,538B
; FILING DATE: 20-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/843,359
; FILING DATE: 25-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,483
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-212399
; FILING DATE: 17-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-212398
; FILING DATE: 17-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-355553
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 62-330219
; FILING DATE: 28-DEC-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30,092
; REFERENCE/DOCKET NUMBER: TEJN-005/02US
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5070
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-963-538B-11
Query Match 100.0%; Score 29; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 1 DDDDK 5
RESULT 11
US-08-668-123-49
; Sequence 49, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-668-123-49
Query Match 100.0%; Score 29; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5

TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-595-868C-16

Query Match 100.0%; Score 29; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 15
US-08-481-435-41
Sequence 41, Application US/08481435
Patent No. 6027906
GENERAL INFORMATION:
APPLICANT: Balganes, Tanjore S
APPLICANT: Town, Christine
TITLE OF INVENTION: No. 6027906el Polypeptides
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,435
FILING DATE: 10-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IN 580/MAS/94
FILING DATE: 01-JUL-1994
PRIOR APPLICATION DATA: SE 9404072-2
FILING DATE: 24-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-151
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-481-435-41

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 16
US-08-483-749A-33
Sequence 33, Application US/08483749A
Patent No. 6054561
GENERAL INFORMATION:
APPLICANT: RING, DAVID B.
TITLE OF INVENTION: ANTIGEN-BINDING SITES OF ANTIBODY
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: INTELLECTUAL PROPERTY - R440, PO BOX 8097
CITY: EMERYVILLE
STATE: CA
COUNTRY: USA
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,749A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SAVERIDE, PAUL B.
REGISTRATION NUMBER: 36,914
REFERENCE/DOCKET NUMBER: 0508.008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2585
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-749A-33

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 17
US-08-997-918-56
Sequence 56, Application US/08997918
Patent No. 6077689
GENERAL INFORMATION:
APPLICANT: Snavelly, Marshall D.
TITLE OF INVENTION: ENHANCED SOLUBILITY OF RECOMBINANT PROTEINS
FILE REFERENCE: A-496
CURRENT APPLICATION NUMBER: US/08/997,918
CURRENT FILING DATE: 1997-12-24
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 56
LENGTH: 5
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

```
; OTHER INFORMATION: Description of Artificial Sequence: Peptidase cut
; OTHER INFORMATION: site
US-09-997-918-56

Query Match          100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 18
US-08-997-532B-11
; Sequence 11, Application US/08997532B
; Patent No. 6087332
; GENERAL INFORMATION:
; APPLICANT: GALLER, Lawrence Isaac
; TITLE OF INVENTION: STREPTOKINASE DERIVATIVES WITH HIGH
; AFFINITY FOR ACTIVATED PLATELETS AND METHODS OF THEIR
; PRODUCTION AND USE IN THROMBOLYTIC THERAPY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,532B
; FILING DATE: 23-DEC-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,884
; REFERENCE/DOCKET NUMBER: GALLER-1
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-997-532B-11

Query Match          100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 19
US-08-888-381-5
; Sequence 5, Application US/08888381
; Patent No. 6110703
; GENERAL INFORMATION:
; APPLICANT: Egel-Mitani, Michi
; APPLICANT: Brandt, Jakob
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: Method For Production of Polypeptides
```

```
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6110703o No. 6110703disk of No. 6110703th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,381
; FILING DATE: 04-June-1999
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4840.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-888-381-5

Query Match          100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 20
US-08-927-128-6
; Sequence 6, Application US/08927128
; Patent No. 6127150
; GENERAL INFORMATION:
; APPLICANT: Coolidge, Thomas
; APPLICANT: Wagner, Fred
; APPLICANT: ven Heeke, Gino
; APPLICANT: Schuster, Sheldon
; APPLICANT: Stout, Jay
; APPLICANT: Wylie, Dwane
; TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 6127150west Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,128
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680,004
```

; FILING DATE: 15-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35, 093
; REFERENCE/DOCKET NUMBER: 8648.2USD1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-927-128-6

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 21

US-09-020-880-37
; Sequence 37, Application US/09020880A
; Patent No. 6136558
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ballinger, Marcus D.
; APPLICANT: Jones, Jennifer T.
; APPLICANT: Fairbrother, Wayne J.
; APPLICANT: Sliwowski, Mark X.
; APPLICANT: Wells, James A.
; TITLE OF INVENTION: HERGULIN VARIANTS
; FILE REFERENCE: 14918-720CON1
; CURRENT APPLICATION NUMBER: US/09/020,880A
; CURRENT FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: US 60/037,581
; EARLIER FILING DATE: 1997-02-10
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 5
; TYPE: PRT
; ORGANISM: No. 6136558 relevant (recombinant)
US-09-020-880-37

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 22

US-09-197-801-15
; Sequence 15, Application US/09197801B
; Patent No. 6159722
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard
; APPLICANT: Hopfner, Karl-Peter
; APPLICANT: Engh, Richard

; APPLICANT: Bode, Wolfram
; APPLICANT: Huber, Robert
; TITLE OF INVENTION: Chimeric Serine Proteases
; FILE REFERENCE: 20119
; CURRENT APPLICATION NUMBER: US/09/197,801B
; CURRENT FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-197-801-15

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 23

US-09-551-028-15
; Sequence 15, Application US/09551028
; Patent No. 6171842
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard
; APPLICANT: Hopfner, Karl-Peter
; APPLICANT: Engh, Richard
; APPLICANT: Bode, Wolfram
; APPLICANT: Huber, Robert
; TITLE OF INVENTION: Chimeric Serine Proteases
; FILE REFERENCE: 20119
; CURRENT APPLICATION NUMBER: US/09/551,028
; CURRENT FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: US/09/197,801
; PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-551-028-15

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 24

US-08-938-595-3
; Sequence 3, Application US/08938595
; Patent No. 6197946
; GENERAL INFORMATION:
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Garner, Ian
; TITLE OF INVENTION: Peptide Production as Fusion Protein in
; TITLE OF INVENTION: Transgenic Mammal Milk
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,595
FILING DATE: (Herewith)
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/727,153
FILING DATE: 08-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/718,523
FILING DATE: 08-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/GB95/00769
FILING DATE: 05-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9406974.7
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fleshner, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0623.0460003
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-938-595-3

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 25
US-08-727-153-3
Sequence 3, Application US/08727153
Patent No. 6211427
GENERAL INFORMATION:
APPLICANT: Cottingham, Ian R.
APPLICANT: Garner, Ian
TITLE OF INVENTION: Peptide Production as Fusion Protein in Transgenic Mammal Milk
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,153
FILING DATE: 08-OCT-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/718,523
FILING DATE: 08-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/GB95/00769
FILING DATE: 05-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9406974.7
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fleshner, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 0623.0460001
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-727-153-3

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 26
US-09-139-819A-16
Sequence 16, Application US/09139819A
Patent No. 6251635
GENERAL INFORMATION:
APPLICANT: WAGNER, Fred W.
APPLICANT: STOUT, Jay S.
APPLICANT: HENRIKSEN, Dennis B.
APPLICANT: FARTRIDGE, Bruce E.
APPLICANT: HOLMQUIST, Bart
APPLICANT: FRANK, Julie A.
TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN AND RELATED ANALOGS
TITLE OF INVENTION: AND RELATED ANALOGS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/139,819A
FILING DATE: 25-AUG-1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/595,868
FILING DATE: 06-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 089187/0144
TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-139-819A-16

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 27

US-08-840-466A-26
Sequence 26, Application US/08840466A
Patent No. 6261561

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/840,466A

FILING DATE: 18-Apr-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laural S.

REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995.0029-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 26:

US-08-840-466A-26

Query Match 100.0%; Score 29; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 28

US-09-260-038B-14
Sequence 14, Application US/09260038B
Patent No. 6348344

GENERAL INFORMATION:

APPLICANT: Maty Aval-Hershkovitz et al.
TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
EXPRESSING RECOMBINANT HEPARANASE
AND METHODS OF PURIFYING SAME

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Caatorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead+ Slimnote-890TX

OPERATING SYSTEM: MS DOS version 6.2,

Windows version 3.11

SOFTWARE: Word for Windows version 2.0 converted to
an ASCII file

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/260,038B

FILING DATE: 02-Mar-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/071,618

FILING DATE: May 1, 1998

APPLICATION NUMBER: 09/071,739

FILING DATE: May 1, 1998

APPLICATION NUMBER: 08/922,180

FILING DATE: September 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 910/16

TELECOMMUNICATION INFORMATION:

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 5

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-260-038B-14

Query Match 100.0%; Score 29; DB 4; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 29

US-09-415-000-3

Sequence 3, Application US/09415000

Patent No. 6379903

GENERAL INFORMATION:

APPLICANT: BRIZZARD, BILLY L.

APPLICANT: HERNAN, RON

TITLE OF INVENTION: PURIFICATION OF RECOMBINANT PROTEINS FUSED TO MULTIPLE

EPITOPES

FILE REFERENCE: SGM6933

```
; CURRENT APPLICATION NUMBER: US/09/415,000
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthesized
; OTHER INFORMATION: sequence
US-09-415-000-3

Query Match      100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DDDDK 5
Db      1 DDDDK 5

RESULT 30
US-09-101-544-37
; Sequence 37, Application US/09101544
; Patent No. 6387638
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ballinger, Marcus D.
; APPLICANT: Jones, Jennifer T.
; APPLICANT: Fairbrother, Wayne J.
; APPLICANT: Sliwkowski, Mark X.
; APPLICANT: Wells, James A.
; TITLE OF INVENTION: HEREGULIN VARIANTS
; FILE REFERENCE: 14918-720CON2
; CURRENT APPLICATION NUMBER: US/09/101,544
; CURRENT FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: US 09/020,880
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: US 60/037,581
; PRIOR FILING DATE: 1997-02-10
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 5
; TYPE: PRT
; ORGANISM: No. 6387638 relevant (recombinant)
US-09-101-544-37

Query Match      100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DDDDK 5
Db      1 DDDDK 5

RESULT 31
US-09-578-303-29
; Sequence 29, Application US/09578303
; Patent No. 6399759
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Whitworth, S. Troy
; APPLICANT: Blum, Murray S.
; TITLE OF INVENTION: Ant Proteases and Methods of Inhibition
; FILE REFERENCE: 235.00150101
; CURRENT APPLICATION NUMBER: US/09/578,303
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,331
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 33
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Trypsin
; OTHER INFORMATION: activation peptide sequence
US-09-578-303-29

Query Match      100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DDDDK 5
Db      1 DDDDK 5

RESULT 32
US-09-696-188B-26
; Sequence 26, Application US/09696188B
; Patent No. 6406885
; GENERAL INFORMATION:
; APPLICANT: Stewart, C. Neal
; APPLICANT: McKee, Marian L.
; APPLICANT: O'Brien, Alison D.
; APPLICANT: Wachtel, Marian R.
; TITLE OF INVENTION: Method Of Stimulating An Immune Response
; By Administration Of Host Organisms That Express Intimin
; Alone Or As A Fusion Protein With One Or More Other
; Antigens.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/696,188B
; FILING DATE: 26-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/840,466
; FILING DATE: 1997-04-18
; ATTORNEY/AGENT INFORMATION:
; NAME: Boone, Laural S.
; REGISTRATION NUMBER: 43,505
; REFERENCE/DOCKET NUMBER: 04995.0029-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-696-188B-26

Query Match      100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 33

US-09-750-913-16
; Sequence 16, Application US/09750913
; Patent No. 6410707

GENERAL INFORMATION:

APPLICANT: WAGNER, Fred W.
STOUT, Jay S.
HENRIKSEN, Dennis B.
PARTRIDGE, Bruce E.
HOLMQUIST, Bart
FRANK, Julie A.

TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
AND RELATED ANALOGS

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/750,913

FILING DATE: 12-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/139,819

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 089187/0144

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 16:

US-09-750-913-16

Query Match 100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 34

US-09-635-923-14

; Sequence 14, Application US/09635923

; Patent No. 6426209

GENERAL INFORMATION:

APPLICANT: Maty Ayal-Hershkovitz et al.

TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
EXPRESSING RECOMBINANT HEPARANASE

AND METHODS OF PURIFYING SAME

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: PC

OPERATING SYSTEM: MS DOS version 6.2,

Windows version 3.11

SOFTWARE: Word for Windows version 6.2,

an ASCII file

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/635,923

FILING DATE: 10-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/487,716

FILING DATE: 19-Jan-2000

APPLICATION NUMBER: 09/071,618

FILING DATE: May 1, 1998

APPLICATION NUMBER: 09/071,739

FILING DATE: May 1, 1998

APPLICATION NUMBER: 08/922,180

FILING DATE: September 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 910/16

TELECOMMUNICATION INFORMATION:

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 5

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-635-923-14

Query Match 100.0%; Score 29; DB 4; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 1 DDDDK 5

RESULT 35

US-07-757-022B-143

; Sequence 143, Application US/07757022B

; Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/57,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserit, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-757-022B-143

Query Match          100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 36
US-09-180-422B-14
; Sequence 14, Application US/09180422B
; Patent No. 6444644
; GENERAL INFORMATION:
; APPLICANT: BRUCKDORFER, KARL R
; ETELAIE, CAMILLE
; TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED
; FROM APOLIPOPROTEIN B-100
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE, P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/180,422B
; FILING DATE: 07-Dec-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SNOFF, B.J.
; REGISTRATION NUMBER: 36663
```

```
;
; REFERENCE/DOCKET NUMBER: 117-268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
; US-09-180-422B-14

Query Match          100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 1 DDDDK 5

RESULT 37
US-09-487-716A-14
; Sequence 14, Application US/09487716A
; Patent No. 6475763
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Herskhovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: PC
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/487,716A
; FILING DATE: 19-Jan-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
; US-09-487-716A-14
```


Query Match 100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 38

US-09-059-625-39
; Sequence 39, Application US/09059625
; Patent No. 6486303
; GENERAL INFORMATION:
; APPLICANT: Moyle, William R.
; TITLE OF INVENTION: Improved Method For Making Hormone
; TITLE OF INVENTION: Heterodimers
; NUMBER OF SEQUENCES: 90
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard R. Muccino
; STREET: 758 Springfield Avenue
; CITY: Summit
; STATE: NJ
; COUNTRY: US
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,625
; FILING DATE: 14-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muccino, Richard R.
; REGISTRATION NUMBER: 32,538
; REFERENCE/DOCKET NUMBER: UMD1-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-273-4988
; TELEFAX: 908-273-4679
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-059-625-39

Query Match 100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 39

US-09-464-152A-6
; Sequence 6, Application US/09464152A
; Patent No. 6498020
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Miroux, Bruno
; TITLE OF INVENTION: Fusion Proteins Comprising Colled-Coil Structures Derived of Bovine
; TITLE OF INVENTION: Atypase Inhibitor Protein
; FILE REFERENCE: 3789/85128
; CURRENT APPLICATION NUMBER: US/09/464,152A

; CURRENT FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: PCT/GB98/02041
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: GB971460.7
; PRIOR FILING DATE: 1997-07-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protease cleavage region targeted by Enterokinase
US-09-464-152A-6

Query Match 100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 40

US-09-185-852-9
; Sequence 9, Application US/09185852
; Patent No. 6537806
; GENERAL INFORMATION:
; APPLICANT: Osborne, William R.A.
; APPLICANT: Ramesh, Nagarajan
; TITLE OF INVENTION: Compositions and Methods for Treating Diabetes
; FILE REFERENCE: P-UW 3264
; CURRENT APPLICATION NUMBER: US/09/185,852
; CURRENT FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/087,660
; EARLIER FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: Sequence
US-09-185-852-9

Query Match 100.0%; Score 29; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 41

PCT-US92-11270-7
; Sequence 7, Application PC/TUS9211270
; GENERAL INFORMATION:
; APPLICANT: Rezaie, Alireza
; APPLICANT: Esmon, Charles T.
; APPLICANT: Morrissey, James H.
; TITLE OF INVENTION: Expression and Purification of
; TITLE OF INVENTION: Recombinant Soluble Tissue Factor
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.

```
;
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11270
; FILING DATE: 19921229
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/816679
; FILING DATE: 03-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/730040
; FILING DATE: 12-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/292447
; FILING DATE: 30-DEC-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/683682
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Cleavage-site
; LOCATION: 1..5
; OTHER INFORMATION: /note= "Enterokinase Cleavage Site"
PCT-US92-11270-7

Query Match 100.0%; Score 29; DB 5; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 DDDDK 5
Db 1 DDDDK 5
```

```
RESULT 42
PCT-US93-06591-2
; Sequence 2, Application PC/TUS9306591
; GENERAL INFORMATION:
; APPLICANT: Stout, Jay
; APPLICANT: Wagner, Fred W.
; APPLICANT: Coolidge, Thomas R.
; APPLICANT: Holmquist, Barton
; TITLE OF INVENTION: METHOD FOR MODIFICATION OF
; RECOMBINANT POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 Northwest Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06591
; FILING DATE: 19930713
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, Albin J.
; REGISTRATION NUMBER: 28,650
; REFERENCE/DOCKET NUMBER: 8648.35-WO-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-06591-2

Query Match 100.0%; Score 29; DB 5; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 DDDDK 5
Db 1 DDDDK 5
```

```
RESULT 43
PCT-US95-15800-2
; Sequence 2, Application PC/TUS9515800
; GENERAL INFORMATION:
; APPLICANT: Biobrasca, Inc.
; TITLE OF INVENTION: PRODUCTION OF PEPTIDES USING
; RECOMBINANT FUSION PROTEIN CONSTRUCTS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 Northwest Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/15800
; FILING DATE: 07-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/350,530
; FILING DATE: 07-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.45USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

TOPLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
PCT-US95-15800-2

Query Match 100.0%; Score 29; DB 5; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 44

US-08-243-082-21
Sequence 21, Application US/08243082
Patent No. 5506120
GENERAL INFORMATION:
APPLICANT: YAMAMOTO, Hiroaki
APPLICANT: YAMASHITA, Kunihiko
TITLE OF INVENTION: METHOD OF PRODUCING PEPTIDES OR
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spencer, Frank & Schneider
STREET: 1111 Nineteenth Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/243,082
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/853,754
FILING DATE: 05-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Schneller, John W.
REGISTRATION NUMBER: 26,031
REFERENCE/DOCKET NUMBER: KUWAT 0010
TELEPHONE: (202) 828-8000
TELEFAX: (202) 828-8038
TELEX: SPENCER 64267
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-243-082-21

Query Match 100.0%; Score 29; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 2 DDDDK 6

RESULT 45

US-08-200-900A-35

Sequence 35, Application US/08200900A
Patent No. 5665566
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,900A
FILING DATE: 23-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meinelert, Maureen C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5201-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-900A-35

Query Match 100.0%; Score 29; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 2 DDDDK 6

Search completed: October 14, 2003, 13:12:00
Job time : 15.1818 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:12:07 ; Search time 20.4545 Seconds
(without alignments)
39.387 Million cell updates/sec

Title: US-09-856-050-19_COPY_36_40
Perfect score: 29
Sequence: 1 DDDDK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues
Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	5	9 US-09-909-652-2	Sequence 2, Appli
2	29	100.0	5	9 US-09-970-308-3	Sequence 3, Appli
3	29	100.0	5	10 US-09-967-386-2	Sequence 2, Appli
4	29	100.0	5	10 US-09-858-332-1	Sequence 1, Appli
5	29	100.0	5	10 US-09-858-332-7	Sequence 7, Appli
6	29	100.0	5	11 US-09-994-487-4	Sequence 4, Appli
7	29	100.0	5	12 US-10-205-110-26	Sequence 26, Appl
8	29	100.0	5	12 US-10-150-058-26	Sequence 26, Appl
9	29	100.0	5	12 US-10-024-597-7	Sequence 7, Appli
10	29	100.0	5	12 US-10-353-908-14	Sequence 14, Appl
11	29	100.0	5	14 US-10-124-557-143	Sequence 143, App
12	29	100.0	5	15 US-10-137-351-14	Sequence 14, Appl
13	29	100.0	5	15 US-10-119-235-4	Sequence 4, Appli
14	29	100.0	5	15 US-10-158-742A-17	Sequence 17, Appl
15	29	100.0	5	15 US-10-274-638-7	Sequence 7, Appli

16	29	100.0	5	15 US-10-328-813-9	Sequence 9, Appli
17	29	100.0	5	15 US-10-023-888-23	Sequence 23, Appl
18	29	100.0	5	16 US-10-082-747A-37	Sequence 37, Appl
19	29	100.0	6	10 US-09-884-767A-8	Sequence 8, Appli
20	29	100.0	6	14 US-10-066-209-7	Sequence 7, Appli
21	29	100.0	6	14 US-10-066-209-9	Sequence 9, Appli
22	29	100.0	7	10 US-09-884-767A-74	Sequence 74, Appl
23	29	100.0	7	10 US-09-884-767A-194	Sequence 194, App
24	29	100.0	7	11 US-09-873-712-30	Sequence 30, Appl
25	29	100.0	7	15 US-10-083-815-2	Sequence 2, Appli
26	29	100.0	8	US-08-971-317A-9	Sequence 9, Appli
27	29	100.0	8	US-08-926-626-12	Sequence 12, Appl
28	29	100.0	8	US-09-754-105-3	Sequence 3, Appli
29	29	100.0	8	9 US-09-050-516-48	Sequence 48, Appl
30	29	100.0	8	9 US-09-276-600-10	Sequence 10, Appl
31	29	100.0	8	9 US-09-771-956-12	Sequence 12, Appl
32	29	100.0	8	9 US-09-785-934-3	Sequence 3, Appli
33	29	100.0	8	9 US-09-835-147-10	Sequence 10, Appl
34	29	100.0	8	9 US-09-760-008A-14	Sequence 14, Appl
35	29	100.0	8	9 US-09-728-911-11	Sequence 11, Appl
36	29	100.0	8	9 US-09-813-329-65	Sequence 65, Appl
37	29	100.0	8	9 US-09-099-823-26	Sequence 26, Appl
38	29	100.0	8	9 US-09-790-264-68	Sequence 68, Appl
39	29	100.0	8	9 US-09-809-517A-8	Sequence 8, Appli
40	29	100.0	8	9 US-09-234-717-24	Sequence 24, Appl
41	29	100.0	8	9 US-09-850-178-18	Sequence 18, Appl
42	29	100.0	8	9 US-09-193-538-22	Sequence 22, Appl
43	29	100.0	8	9 US-09-250-883-22	Sequence 22, Appl
44	29	100.0	8	9 US-09-735-368-4	Sequence 4, Appli
45	29	100.0	8	9 US-09-096-259-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-09-909-652-2
; Sequence 2, Application US/0909652
; Patent No. US20020025537A1
; GENERAL INFORMATION:
; APPLICANT: Kairos Scientific, Inc.
; APPLICANT: Bylina, Edward J.
; APPLICANT: Coleman, William J.
; APPLICANT: Youvan, Douglas C.
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR GENERATING
; FILE REFERENCE: 22346-7001
; CURRENT APPLICATION NUMBER: US/09/909,652
; PRIOR FILING DATE: 2001-10-15
; PRIOR FILING DATE: US 60/219,179
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enterokinase recognition site
US-09-909-652-2

Query Match 100.0%; Score 29; DB 9; Length 5;
Best Local Similarity 100.0%; Pred.No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db' 1 DDDDK 5

RESULT 2

US-09-970-308-3
; Sequence 3, Application US/09970308

```
; Patent No. US20020045193A1
; GENERAL INFORMATION:
; APPLICANT: BRIZZARD, BILLY L.
; APPLICANT: HERNAN, RON
; TITLE OF INVENTION: PURIFICATION OF RECOMBINANT PROTEINS FUSED TO MULTIPLE
; TITLE OF INVENTION: EPITOPES
; FILE REFERENCE: SGM 6933.2
; CURRENT APPLICATION NUMBER: US/09/970,308
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 09/415,000
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthesized
; OTHER INFORMATION: sequence
US-09-970-308-3

Query Match      100.0%; Score 29; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 3
US-09-967-386-2
; Sequence 2, Application US/09967386
; Patent No. US2002015992A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Henkin, Jack
; APPLICANT: Davidson, Donald J.
; TITLE OF INVENTION: ANTIANGIOGENIC POLYPEPTIDES AND METHODS
; TITLE OF INVENTION: FOR INHIBITING ANGIOGENESIS
; FILE REFERENCE: 6738.US.02
; CURRENT APPLICATION NUMBER: US/09/967,386
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/236,550
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enterokinase Cleavage Site
US-09-967-386-2

Query Match      100.0%; Score 29; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 4
US-09-858-332-1
; Sequence 1, Application US/09858332
; Patent No. US20020164718A1
; GENERAL INFORMATION:
; APPLICANT: Tchaga, Grigory S.
; APPLICANT: Jokhadze, George
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for
```

```
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: CLON056CIP
; CURRENT APPLICATION NUMBER: US/09/858,332
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: 09/404,017
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,867
; PRIOR FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-858-332-1

Query Match      100.0%; Score 29; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 5
US-09-858-332-7
; Sequence 7, Application US/09858332
; Patent No. US20020164718A1
; GENERAL INFORMATION:
; APPLICANT: Tchaga, Grigory S.
; APPLICANT: Jokhadze, George
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: CLON056CIP
; CURRENT APPLICATION NUMBER: US/09/858,332
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: 09/404,017
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,867
; PRIOR FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-858-332-7

Query Match      100.0%; Score 29; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 6
US-09-994-487-4
; Sequence 4, Application US/09994487
; Publication No. US20030099954A1
; GENERAL INFORMATION:
; APPLICANT: Miltenyi, Stefan
; APPLICANT: Kohler, Matthias
; TITLE OF INVENTION: Apparatus and method for modification of
; TITLE OF INVENTION: magnetically immobilized biomolecules
; FILE REFERENCE: MILT0004
; CURRENT APPLICATION NUMBER: US/09/994,487
```

; CURRENT FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-994-487-4

Query Match 100.0%; Score 29; DB 11; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 7

US-10-205-110-26
; Sequence 26, Application US/10205110
; Publication No. US20030144471A1
; GENERAL INFORMATION:
; APPLICANT: Jonassen, Ib
; APPLICANT: Egel-Mitani, Michi
; APPLICANT: Balschmidt, Per
; APPLICANT: Markussen, Jan
; APPLICANT: Diers, Ivan
; APPLICANT: Kjeldsen, Thomas Borglum
; TITLE OF INVENTION: Method for Making Acylated Polypeptides
; FILE REFERENCE: 6289.200-US
; CURRENT APPLICATION NUMBER: US/10/205,110
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: PA 2001 01141
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/310,793
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-205-110-26

Query Match 100.0%; Score 29; DB 12; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 8

US-10-150-058-26
; Sequence 26, Application US/10150058
; Publication No. US20030147902A1
; GENERAL INFORMATION:
; APPLICANT: Stewart, C. Neal
; Mckee, Marian L.
; O'Brien, Alison D.
; Wachtel, Marian R.

; TITLE OF INVENTION: Method Of Stimulating An Immune Response
; By Administration Of Host Organisms That Express Intimin
; Alone Or As A Fusion Protein With One Or More Other
; Antigens.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/150,058
; FILING DATE: 20-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,466A
; FILING DATE: 18-Apr-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Boone, Laural S.
; REGISTRATION NUMBER: 43,505
; REFERENCE/DOCKET NUMBER: 04995.0029-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-150-058-26

Query Match 100.0%; Score 29; DB 12; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
| | | | |
Db 1 DDDDK 5

RESULT 9

US-10-024-597-7
; Sequence 7, Application US/10024597
; Publication No. US20030167477A1
; GENERAL INFORMATION:
; APPLICANT: Cottingham, Ian R.
; APPLICANT: McCreath, Graham E.
; TITLE OF INVENTION: Fusion Proteins Incorporating Lysozyme
; FILE REFERENCE: 0623.0730002/EKS/BJD
; CURRENT APPLICATION NUMBER: US/10/024,597
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US (to be assigned)
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/GB00/02459
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: GB 9914733.2
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: US 60/147,819
; PRIOR FILING DATE: 1999-08-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Enterokinase
; OTHER INFORMATION: cleavage site

US-10-024-597-7

Query Match 100.0%; Score 29; DB 12; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 10

US-10-353-908-14
; Sequence 14, Application US/10353908
; Publication No. US20030170770A1
; GENERAL INFORMATION:
; APPLICANT: KHANNA, PYARE
; APPLICANT: FUNG, PETER
; APPLICANT: HORECKA, JOSEPH L.
; TITLE OF INVENTION: ENZYME ACTIVATION PROTEASE ASSAY
; FILE REFERENCE: 3817.09-1
; CURRENT APPLICATION NUMBER: US/10/353,908
; CURRENT FILING DATE: 2003-01-28
; PRIOR APPLICATION NUMBER: 60/352,780
; PRIOR FILING DATE: 2002-01-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-10-353-908-14

Query Match 100.0%; Score 29; DB 12; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 11

US-10-124-557-143
; Sequence 143, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesser, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Cserti, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 143:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
TOPOLOGY: unknown

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-10-124-557-143

Query Match 100.0%; Score 29; DB 14; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 12

US-10-137-351-14
; Sequence 14, Application US/10137351
; Publication No. US20030068806A1
; GENERAL INFORMATION:
; APPLICANT: Ayal-Hershkovitz, Maty
; APPLICANT: Pecker, Iris
; APPLICANT: Yacoby-Zeevi, Oron
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR EXPRESSING RECOMBINANT
; FILE REFERENCE: 02/23665
; CURRENT APPLICATION NUMBER: US/10/137,351
; CURRENT FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Enterokinase recognition and cleavage site
US-10-137-351-14

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 13

US-10-119-235-4
; Sequence 4, Application US/10119235
; Publication No. US20030096760A1
; GENERAL INFORMATION:
; APPLICANT: Holt, Dennis A.
; APPLICANT: Veber, Daniel F.

; APPLICANT: Yamashita, Dennis S.
; TITLE OF INVENTION: Method of Antagonizing the Human SRC SH2
; FILE REFERENCE: P50630
; CURRENT APPLICATION NUMBER: US/10/119,235
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 1998-09-02
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/040,658
; PRIOR FILING DATE: 1997-03-10
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 5
; TYPE: PRT
; ORGANISM: homo sapien
US-10-119-235-4

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 14
US-10-158-742A-17
; Sequence 17, Application US/10158742A
; Publication No. US20030104581A1
; GENERAL INFORMATION:
; APPLICANT: Hoess, Eva
; APPLICANT: Meier, Thomas
; APPLICANT: Pestlin, Gabriele
; APPLICANT: Poppp, Friedrich
; APPLICANT: Reichert, Klaus
; APPLICANT: Schmuck, Rainer
; APPLICANT: Schneider, Bernd
; APPLICANT: Seidel, Christoph
; APPLICANT: Tischer, Wilhelm
; TITLE OF INVENTION: PROCESS FOR MAKING ANTIFUSOGENIC FUSION PEPTIDES THAT FORM
; FILE REFERENCE: 20904
; CURRENT APPLICATION NUMBER: US/10/158,742A
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: EP 01114497.9
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE:
; SEQ ID NO 17
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cleavage
; OTHER INFORMATION: sequence
US-10-158-742A-17

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 15
US-10-274-638-7
; Sequence 7, Application US/10274638
; Publication No. US20030109000A1
; GENERAL INFORMATION:

; APPLICANT: Moore, Margaret D.
; APPLICANT: Fox, Brian A.
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS
; FILE REFERENCE: AND METHODS FOR PRODUCING IT
; FILE REFERENCE: 01-30
; CURRENT APPLICATION NUMBER: US/10/274,638
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: 60/346,117
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: polypeptide, enterokinase cleavage site
US-10-274-638-7

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 16
US-10-328-813-9
; Sequence 9, Application US/10328813
; Publication No. US20030113305A1
; GENERAL INFORMATION:
; APPLICANT: Osborne, William R.A.
; APPLICANT: Ramesh, Nagarajan
; TITLE OF INVENTION: Compositions and Methods for Treating Diabetes
; FILE REFERENCE: P-UW 3264
; CURRENT APPLICATION NUMBER: US/10/328,813
; CURRENT FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US/09/185,852
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: 60/087,660
; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: Sequence
US-10-328-813-9

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 1 DDDDK 5

RESULT 17
US-10-023-888-23
; Sequence 23, Application US/10023888
; Publication No. US20030119088A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE
; FILE REFERENCE: 2035150577
; CURRENT APPLICATION NUMBER: US/10/023,888
; CURRENT FILING DATE: 2001-12-21

; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-023-888-23

Query Match 100.0%; Score 29; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 18
US-10-082-747A-37
; Sequence 37, Application US/10082747A
; Publication No. US20030129688A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ballinger, Marcus D.
; APPLICANT: Jones, Jennifer T.
; APPLICANT: Fairbrother, Wayne J.
; APPLICANT: Sliwkowski, Mark X.
; APPLICANT: Wells, James A.
; TITLE OF INVENTION: HERGULIN VARIANTS
; FILE REFERENCE: 402E-476112US
; CURRENT APPLICATION NUMBER: US/10/082.747A
; CURRENT FILING DATE: 2002-09-16
; PRIOR FILING DATE: US 09/101,544
; PRIOR APPLICATION NUMBER: PCT/US/98/01579
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: US 08/799,054
; PRIOR FILING DATE: 1997-02-10
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Enterokinase protease recognition site
US-10-082-747A-37

Query Match 100.0%; Score 29; DB 16; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 19
US-09-884-767A-8
; Sequence 8, Application US/09884767A
; Publication No. US20020192789A1
; GENERAL INFORMATION:
; APPLICANT: DYAX Corp.
; APPLICANT: Lev, Arthur C.
; APPLICANT: Luneau, Christopher J.
; APPLICANT: Ladner, Robert C
; TITLE OF INVENTION: NOVEL ENTEROKINASE CLEAVAGE SEQUENCES
; FILE REFERENCE: DYX-012.1 US, DYX-012.1 PCT
; CURRENT APPLICATION NUMBER: US/09/884,767A
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/597,321

; PRIOR FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic enterokinase cleavage sequence
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: Xaa is Ile (natural trypsinogen site) or any amino acid (synth
; OTHER INFORMATION: C cleavage sites)
US-09-884-767A-8

Query Match 100.0%; Score 29; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 1 DDDDK 5

RESULT 20
US-10-066-209-7
; Sequence 7, Application US/10066209
; Publication No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: GH-50030-D1
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 6
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-7

Query Match 100.0%; Score 29; DB 14; Length 6;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
|||||
Db 2 DDDDK 6

RESULT 21
US-10-066-209-9
; Sequence 9, Application US/10066209
; Publication No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: GH-50030-D1
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25

; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 6
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-9

Query Match 100.0%; Score 29; DB 14; Length 6;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 2 DDDDK 6

RESULT 22
US-09-884-767A-74
; Sequence 74, Application US/09884767A
; Publication No. US20020192789A1
; GENERAL INFORMATION:
; APPLICANT: DYAX Corp.
; APPLICANT: Ley, Arthur C.
; APPLICANT: Luneau, Christopher J.
; APPLICANT: Ladner, Robert C.
; TITLE OF INVENTION: NOVEL ENTEROKINASE CLEAVAGE SEQUENCES
; FILE REFERENCE: DYX-012.1 US, DYX-012.1 PCT
; CURRENT APPLICATION NUMBER: US/09/884,767A
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/597,321
; PRIOR FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: natural enterokinase cleavage sequence
US-09-884-767A-74

Query Match 100.0%; Score 29; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 2 DDDDK 6

RESULT 23
US-09-884-767A-194
; Sequence 194, Application US/09884767A
; Publication No. US20020192789A1
; GENERAL INFORMATION:
; APPLICANT: DYAX Corp.
; APPLICANT: Ley, Arthur C.
; APPLICANT: Luneau, Christopher J.
; APPLICANT: Ladner, Robert C.
; TITLE OF INVENTION: NOVEL ENTEROKINASE CLEAVAGE SEQUENCES
; FILE REFERENCE: DYX-012.1 US, DYX-012.1 PCT
; CURRENT APPLICATION NUMBER: US/09/884,767A
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/597,321

; PRIOR FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 194
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic enterokinase cleavage sequence
US-09-884-767A-194

Query Match 100.0%; Score 29; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 2 DDDDK 6

RESULT 24
US-09-872-712-30
; Sequence 30, Application US/09872712
; Publication No. US20030059461A1
; GENERAL INFORMATION:
; APPLICANT: Backer, Marina V.
; APPLICANT: Backer, Joseph M.
; TITLE OF INVENTION: MOLECULAR DELIVERY VEHICLE FOR DELIVERY
; TITLE OF INVENTION: OF SELECTED COMPOUNDS TO TARGETS
; FILE REFERENCE: 102131-200
; CURRENT APPLICATION NUMBER: US/09/872,712
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 60/209,660
; PRIOR FILING DATE: 2000-06-05
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Bovine
US-09-872-712-30

Query Match 100.0%; Score 29; DB 11; Length 7;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 3 DDDDK 7

RESULT 25
US-10-083-815-2
; Sequence 2, Application US/10083815
; Publication No. US20030026781A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, Christen M.
; APPLICANT: Clevenger, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; TITLE OF INVENTION: ENDOGENOUS INHIBITOR OF ATP SYNTHASE, INCLUDING
; TITLE OF INVENTION: TREATMENT FOR DIABETES
; FILE REFERENCE: 660088.435C2
; CURRENT APPLICATION NUMBER: US/10/083,815
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Epitope tag
US-10-083-815-2

```
Query Match      100.0%; Score 29; DB 15; Length 7;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DDDDK 5
Db      3 DDDDK 7

RESULT 26
US-08-971-317A-9
; Sequence 9, Application US/08971317A
; Publication No. US20010010925A1
; GENERAL INFORMATION:
; APPLICANT: Willey, Steven R.
; TITLE OF INVENTION: TNF-DELTA LIGAND AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971.317A
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goller, Mimi C
; REGISTRATION NUMBER: 39,046
; REFERENCE/DOCKET NUMBER: 6255.US.01
; TELEPHONE: (847) 935-7550
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. US20010010925A1e
US-08-971-317A-9

Query Match      100.0%; Score 29; DB 8; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DDDDK 5
Db      4 DDDDK 8

RESULT 27
US-08-926-626-12
; Sequence 12, Application US/08926626
; Publication No. US20020035244A1
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.

Query Match      100.0%; Score 29; DB 8; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DDDDK 5
Db      4 DDDDK 8

RESULT 28
US-09-754-105-3
; Sequence 3, Application US/09754105
; Patent No. US20010009768A1
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas
; APPLICANT: Reddy, Pranhitha
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING CYTOKINE DESIGNATED LERK-5
; FILE REFERENCE: 28232
; CURRENT APPLICATION NUMBER: US/09/754,105
; CURRENT FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 09/329,531
; PRIOR FILING DATE: 1999-06-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Flag peptide
; NAME/KEY: Artificial
; LOCATION: (1)..(8)
```

US-09-754-105-3

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 29

US-09-050-516-48
; Sequence 48, Application US/09050516
; Patent No. US20010010904A1

; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: GORDON, JULIAN

; APPLICANT: GRANADOS, EDWARD N.

; APPLICANT: HAYDEN, MARK

; APPLICANT: HODGES, STEVEN C.

; APPLICANT: KLASS, MICHAEL R.

; APPLICANT: KRATOCHVIL, JON D.

; APPLICANT: ROBERTS-RAPP, LISA

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STROUPE, STEPHEN D.

; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE GASTROINTESTINAL

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/050,516

; FILING DATE:
; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,855

; FILING DATE: 31-MAR-1997

; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6065, US.P1

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:
; INFORMATION FOR SEQ ID NO: 48:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-050-516-48

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 4 DDDDK 8
|||||

RESULT 30

US-09-276-600-10

; Sequence 10, Application US/09276600
; Patent No. US20010010908A1

; GENERAL INFORMATION:
; APPLICANT: Patricia Billing-Medel

; APPLICANT: Maurice Cohen

; APPLICANT: Tracey L. Colpitts

; APPLICANT: Julian Gordon

; APPLICANT: Edward N. Granados

; APPLICANT: John C. Russell

; APPLICANT: Stephen D. Stroupe

; TITLE OF INVENTION: Reagents and Methods Useful for
; TITLE OF INVENTION: Detecting Disease of the Prostate

; FILE REFERENCE: 6397.US.01

; CURRENT APPLICATION NUMBER: US/09/276,600

; CURRENT FILING DATE: 1999-03-25

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 10

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Affinity purification system recognition site

US-09-276-600-10

Query Match 100.0%; Score 29; DB 9; Length 8;

Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 4 DDDDK 8

|||||

RESULT 31

US-09-771-956-12

; Sequence 12, Application US/09771956
; Patent No. US20010031474A1

; GENERAL INFORMATION:
; APPLICANT: Bennett, Michele

; APPLICANT: Brodbeck, Robbin

; APPLICANT: Krause, James

; TITLE OF INVENTION: Chimeric Neuropeptide Y Receptors

; FILE REFERENCE: N2000.001

; CURRENT APPLICATION NUMBER: US/09/771,956

; CURRENT FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 12

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: FLAG EPITOPE

US-09-771-956-12

Query Match 100.0%; Score 29; DB 9; Length 8;

Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

Db 4 DDDDK 8

|||||

RESULT 32

```
US-09-785-934-3
; Sequence 3, Application US/09785934
; Patent No. US20020002132A1
; GENERAL INFORMATION:
; APPLICANT: Pluenneke, John
; TITLE OF INVENTION: USE OF INTERLEUKIN-4 ANTAGONISTS AND COMPOSITIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 3005
; CURRENT APPLICATION NUMBER: US/09/785,934
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 09/579,808
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: FLAG Peptide
US-09-785-934-3
Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 4 DDDDK 8
RESULT 33
US-09-835-147-10
; Sequence 10, Application US/09835147
; Patent No. US20020002277A1
; GENERAL INFORMATION:
; APPLICANT: Waliszewski, Charles R.
; APPLICANT: Gayle III, Richard B.
; APPLICANT: Price, Virginia L.
; APPLICANT: Gimpel, Steven D.
; TITLE OF INVENTION: Inhibitors of Platelet Activation and Recruitment
; FILE REFERENCE: 2879-US
; CURRENT APPLICATION NUMBER: US/09/835,147
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/104,585
; PRIOR FILING DATE: 1998-10-16
; PRIOR APPLICATION NUMBER: US 60/107,466
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: US 60/149,010
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: PCT/US99/22955
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-835-147-10
Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 4 DDDDK 8
RESULT 34
US-09-785-934-3
; Sequence 3, Application US/09785934
; Patent No. US20020002132A1
; GENERAL INFORMATION:
; APPLICANT: Pluenneke, John
; TITLE OF INVENTION: USE OF INTERLEUKIN-4 ANTAGONISTS AND COMPOSITIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 3005
; CURRENT APPLICATION NUMBER: US/09/785,934
; CURRENT FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 09/579,808
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: FLAG Peptide
US-09-785-934-3
Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 4 DDDDK 8
RESULT 35
US-09-728-911-11
; Sequence 11, Application US/09728911
; Patent No. US20020012669A1
; GENERAL INFORMATION:
; APPLICANT: Presnell, Scott R.
; APPLICANT: Xu, Wenfeng
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Chen, Zhi
; TITLE OF INVENTION: Human Cytokine Receptor
; FILE REFERENCE: 99-93
; CURRENT APPLICATION NUMBER: US/09/728,911
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/169,049
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: US 60/232,219
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: US 60/244,610
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: FLAG Tag amino acid sequence
US-09-728-911-11
US-09-760-008A-14
; Sequence 14, Application US/09760008A
; Patent No. US20020004483A1
; GENERAL INFORMATION:
; APPLICANT: NISSEN, TORBEN LAUESGAARD
; APPLICANT: ANDERSEN, KIM VILBOUR
; APPLICANT: HANSEN, CHRISTIAN KARSTEN
; APPLICANT: MIKKELSEN, JAN MOLLER
; TITLE OF INVENTION: G-CSF CONJUGATES
; FILE REFERENCE: 31-000700US
; CURRENT APPLICATION NUMBER: US/09/760,008A
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/176,376
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/189,506
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 60/215,644
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DK PA 2000 00024
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: DK PA 2000 00341
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: DK PA 2000 00943
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-760-008A-14
Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 4 DDDDK 8
```

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 36

US-09-813-329-65
; Sequence 65, Application US/09813329
; Patent No. US20020012968A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: No. US20020012968A1el Drosophila Tumor Necrosis Factor Class Mole
; FILE REFERENCE: D0016.np
; CURRENT APPLICATION NUMBER: US/09/813,329
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,816
; PRIOR FILING DATE: 2000-03-21
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 65
; LENGTH: 8
; TYPE: PRT
; ORGANISM: bacteriophage T7
US-09-813-329-65

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 37

US-09-099-823-26
; Sequence 26, Application US/09099823
; Patent No. US20020018990A1
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN MAURICE
; APPLICANT: COLPITS, TRACEY L.
; APPLICANT: FRIEDMAN, JULIAN
; APPLICANT: GORDON, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: SCHEFFEL, CHRISTI
; APPLICANT: STROUPE, STEPHEN D.
; APPLICANT: YU, HONG
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE BREAST
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/099,823
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/879,354
; APPLICATION NUMBER: 20-JUN-1997
; FILING DATE: 20-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6120.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. US20020018990A1e
US-09-099-823-26

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 38

US-09-790-264-68
; Sequence 68, Application US/09790264
; Patent No. US20020028508A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNASTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: 07334-322001
; CURRENT APPLICATION NUMBER: US/09/790,264
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 09/065,661
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/298,531
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: US 09/065,363
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/337,930
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/102,705
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: US 09/363,630
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 09/124,538
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: flag epitope
US-09-790-264-68

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 4 DDDDK 8

RESULT 39
US-09-809-517A-8
; Sequence 8, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1el methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-8

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 4 DDDDK 8

RESULT 40
US-09-234-717-24
; Sequence 24, Application US/09234717A
; Patent No. US20020034740A1
; GENERAL INFORMATION:
; APPLICANT: Patricia Billing-Medel
; APPLICANT: Maurice Cohen
; APPLICANT: Tracey L. Colpitts
; APPLICANT: Paula N. Friedman
; APPLICANT: Julian Gordon
; APPLICANT: Edward N. Granados
; APPLICANT: Steven C. Hodges
; APPLICANT: Michael R. Klass
; APPLICANT: Jon D. Kratochvil
; APPLICANT: John C. Russell
; APPLICANT: Stephen D. Stroupe
; TITLE OF INVENTION: Reagents and Methods Useful for Detecting Diseases of
; FILE REFERENCE: 6450.US.01
; CURRENT APPLICATION NUMBER: US/09/234,717A
; CURRENT FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Affinity purification system recognition site
US-09-234-717-24

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
 |||||
Db 4 DDDDK 8

RESULT 41
US-09-850-178-18
; Sequence 18, Application US/09850178
; Patent No. US20020034749A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Billing-Medel, Patricia A.
; APPLICANT: Cohen, Maurice
; APPLICANT: Colpitts, Tracey L.
; APPLICANT: Friedman, Paula N.
; APPLICANT: Russell, John C.
; APPLICANT: Granados, Edward N.
; APPLICANT: Hodges, Steven C.
; APPLICANT: Klass, Michael R.
; APPLICANT: Kratochvil, Jon D.
; APPLICANT: Roberts-Rapp, Lisa
; APPLICANT: Stroupe, Stephen D.
; APPLICANT: Gordon, Julian
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
; FILE REFERENCE: 6251.US.P1
; CURRENT APPLICATION NUMBER: US/09/850,178
; CURRENT FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: US 08/972,376
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Affinity purification system recognition site
US-09-850-178-18

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
Db 4 DDDDK 8

RESULT 42
US-09-193-538-22
; Sequence 22, Application US/09193538A
; Patent No. US20020037503A1
; GENERAL INFORMATION:
; APPLICANT: Patricia Billing-Medel
; APPLICANT: Maurice Cohen
; APPLICANT: Tracey L. Colpitts
; APPLICANT: Paula N. Friedman
; APPLICANT: Julian Gordon
; APPLICANT: Edward N. Granados
; APPLICANT: Steven C. Hodges
; APPLICANT: Michael R. Klass
; APPLICANT: Jon D. Kratochvil
; APPLICANT: Lisa Roberts-Rapp
; APPLICANT: John C. Russell
; APPLICANT: Stephen D. Stroupe
; TITLE OF INVENTION: Reagents and Methods Useful for Detecting Diseases of the
; FILE REFERENCE: 6193.US.P1
; CURRENT APPLICATION NUMBER: US/09/193,538A
; CURRENT FILING DATE: 1998-11-17

;; EARLIER APPLICATION NUMBER: US 08/971,772
;; EARLIER FILING DATE: 1997-11-17
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 22
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Affinity purification system recognition site
US-09-193-538-22

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 43
US-09-250-883-22
;; Sequence 22, Application US/09250883
;; Patent No. US20020042049A1
;; GENERAL INFORMATION:
;; APPLICANT: Russell, John
;; APPLICANT: Colpitts, Tracey

;; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
;; FOR DETECTING DISEASE OF THE BREAST
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Abbott Laboratories
;; STREET: 100 Abbott Park Road
;; CITY: Abbott Park
;; STATE: IL
;; COUNTRY: USA

;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/250,883
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/889,316
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Becker, Cheryl L

;; REGISTRATION NUMBER: 35,441
;; REFERENCE/DOCKET NUMBER: 6131.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 847/935-1729
;; TELEFAX: 847/938-2623
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 22:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-09-250-883-22

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 44
US-09-735-368-4
;; Sequence 4, Application US/09735368
;; Patent No. US20020042094A1
;; GENERAL INFORMATION:
;; APPLICANT: Venezia, Domenick R.
;; APPLICANT: Taft, David W.
;; APPLICANT: Whitmore, Theodore E.
;; TITLE OF INVENTION: RING FINGER PROTEIN ZAPOF2
;; FILE REFERENCE: 98-11
;; CURRENT APPLICATION NUMBER: US/09/735,368
;; CURRENT FILING DATE: 2000-12-12
;; PRIOR APPLICATION NUMBER: US 60/171,258
;; PRIOR FILING DATE: 1999-12-15
;; NUMBER OF SEQ ID NOS: 11
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 4
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Flag peptide sequence
US-09-735-368-4

Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 4 DDDDK 8

RESULT 45
US-09-096-259-32
;; Sequence 32, Application US/09096259
;; Patent No. US20020045164A1
;; GENERAL INFORMATION:
;; APPLICANT: BILLING-MEDEL, PATRICIA
;; APPLICANT: COHEN, MAURICE
;; APPLICANT: COLPITTS, TRACEY L.
;; APPLICANT: FRIEDMAN, PAULA N.
;; APPLICANT: GORDON, JULIAN
;; APPLICANT: GRANADOS, EDWARD N.
;; APPLICANT: HODGES, STEVEN C.
;; APPLICANT: KLASS, MICHAEL R.
;; APPLICANT: KRATOCHVIL, JON D.
;; APPLICANT: ROBERTS-RAPP, LISA
;; APPLICANT: RUSSELL, JOHN C.
;; APPLICANT: STROUPE, STEPHEN D.
;; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
;; FOR DETECTING DISEASES OF THE LUNG
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Abbott Laboratories
;; STREET: 100 Abbott Park Road
;; CITY: Abbott Park
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60064-3500
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/096,259
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/049,183

```
; FILING DATE: 11-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6113.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-096-259-32
```

```
Query Match 100.0%; Score 29; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 4 DDDDK 8
```

```
Search completed: October 14, 2003, 13:16:20
Job time : 22.4545 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 11.8182 Seconds
(without alignments)
40.687 Million cell updates/sec

Title: US-09-856-050-19_COPY_36_40

Perfect score: 29

Sequence: 1 DDDDK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	29	100.0	8	2 A61328	trypsin (EC 3.4.21
2	29	100.0	20	2 S50022	trypsin-like prote
3	29	100.0	40	2 S50020	trypsin-like prote
4	29	100.0	47	2 T10471	potassium channel -
5	29	100.0	49	2 B26815	peptide PVL-amide
6	29	100.0	92	2 T20731	hypothetical prote
7	29	100.0	109	2 T21867	hypothetical prote
8	29	100.0	140	2 T32477	hypothetical prote
9	29	100.0	141	2 T33710	hypothetical prote
10	29	100.0	144	2 T23275	hypothetical prote
11	29	100.0	165	2 A36237	melanin-concentrat
12	29	100.0	166	2 T33056	hypothetical prote
13	29	100.0	177	1 R8BYGI	regulatory protein
14	29	100.0	185	2 D85040	hypothetical prote
15	29	100.0	198	2 A96520	hypothetical prote
16	29	100.0	212	2 F84481	Mutator-like trans
17	29	100.0	213	2 T37039	hypothetical prote
18	29	100.0	229	1 TRBOTR	trypsin (EC 3.4.21
19	29	100.0	229	1 TRDFS	trypsin (EC 3.4.21
20	29	100.0	231	1 TRPCTR	trypsin (EC 3.4.21
21	29	100.0	232	2 T22698	hypothetical prote
22	29	100.0	239	2 T51270	hypothetical prote
23	29	100.0	246	1 TRRT2	trypsin (EC 3.4.21
24	29	100.0	246	1 TRDGC	trypsin (EC 3.4.21
25	29	100.0	246	2 B25528	trypsin (EC 3.4.21
26	29	100.0	247	1 A25852	trypsin (EC 3.4.21
27	29	100.0	247	1 B25852	trypsin (EC 3.4.21
28	29	100.0	247	1 TRDGS	trypsin (EC 3.4.21
29	29	100.0	247	2 A27547	trypsin (EC 3.4.21

30 29 100.0 247 2 S13813 trypsin (EC 3.4.21

31 29 100.0 247 2 S12764 trypsin (EC 3.4.21

32 29 100.0 250 2 B84708 probable signal pe

33 29 100.0 251 2 T45613 SNAP25-like protei

34 29 100.0 253 2 B29653 aspartic acid-rich

35 29 100.0 259 2 T38363 trypsin (EC 3.4.21

36 29 100.0 263 2 A86272 protein F16A14.10

37 29 100.0 271 2 S64615 hypothetical prote

38 29 100.0 288 2 A86384 unknown protein [i

39 29 100.0 288 2 B40722 homeotic protein 1

40 29 100.0 297 2 S64308 hypothetical prote

41 29 100.0 298 2 T37251 homeobox protein c

42 29 100.0 304 2 S33496 trypsin (EC 3.4.21

43 29 100.0 310 2 T37437 probable 34.9K pro

44 29 100.0 310 2 T31799 hypothetical prote

45 29 100.0 315 2 T34409 hypothetical prote

ALIGNMENTS

RESULT 1

A61328

trypsin (EC 3.4.21.4) precursor - minke whale (fragment)

C;Species: Balaenoptera acutorostrata (minke whale, lesser rorqual)

C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999

C;Accession: A61328

R;Briceux-Gregoire, S.; Schyns, R.; Florquin, M.; Emmens, M.; Wellings, G.W.; Beintem-

Biochim. Biophys. Acta 386, 244-255, 1975

A;Title: N-terminal amino acid sequence of trypsinogen from the lesser rorqual, Bala-

creas.

A;Reference number: A61328; MUID:75146765; PMID:1125273

A;Accession: A61328

A;Molecule type: protein

A;Residues: 1-8

C;Keywords: hydrolase; protein digestion; serine proteinase; zymogen

F;1-8/Domain: activation peptide #status experimental <APT>

Query Match 100.0%; Score 29; DB 2; Length 8;

Best Local Similarity 100.0%; Pred. No. 2.8e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5

Db 4 DDDDK 8

RESULT 2

S50022

trypsin-like protein - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999

C;Accession: S50022

R;Koshikawa, N.; Yasumitsu, H.; Nagashima, Y.; Umeda, M.; Miyazaki, K.

Biochem. J. 303, 187-190, 1994

A;Title: Identification of one- and two-chain forms of trypsinogen 1 produced by a h

A;Reference number: S50020; MUID:95031957; PMID:7945238

A;Accession: S50022

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-20 <KOS>

C;Superfamily: trypsin; trypsin homology

Query Match 100.0%; Score 29; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5

Db 4 DDDDK 8

RESULT 3

S50020
trypsin-like protein - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
C:Accession: S50020
R:Koshikawa, N.; Yasumitsu, H.; Nagashima, Y.; Umeda, M.; Miyazaki, K.
Biochem. J. 303: 187-190, 1994
A:Title: Identification of one- and two-chain forms of trypsinogen 1 produced by a human
A:Reference number: S50020; MUID:95031957; PMID:7945238
A:Accession: S50020
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-20;21-40 <KOS>
C:Superfamily: trypsin; trypsin homology

Query Match 100.0%; Score 29; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 4 DDDDK 8

RESULT 4
T10471
potassium channel-blocking toxin K-alpha - Pandinus imperator (fragment)
C:Species: Pandinus imperator
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T10471
R:Rogowski, R.S.; Collins, J.H.; O'Neill, T.J.; Gustafson, T.A.; Werkman, T.R.; Rogowski, M.
Mol. Pharmacol. 50, 1167-1177, 1996
A:Title: Three new toxins from the scorpion Pandinus imperator selectively block certain
A:Reference number: 217036; MUID:97070422; PMID:8913348
A:Accession: T10471
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-47 <ROG>
A:Cross-references: EMBL:U79579; NID:g1945405; PID:g1945406
C:Genetics:
A:Gene: PTX-1
C:Superfamily: kallotoxin
C:Keywords: neurotoxin; potassium channel inhibitor; venom

Query Match 100.0%; Score 29; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 8 DDDDK 12

RESULT 5
B26815
peptide PYL-amide 2 precursor - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 03-Nov-1987 #sequence_revision 03-Nov-1987 #text_change 18-Aug-2000
C:Accession: B26815
R:Hoffmann, W.; Richter, K.; Kreil, G.
EMBO J. 2, 711-714, 1983
A:Title: A novel peptide designated PYLa and its precursor as predicted from cloned mRNA
A:Reference number: A26815; MUID:84057748; PMID:6688991
A:Accession: B26815
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-49 <HOF>
C:Superfamily: xenopsin precursor

Query Match 100.0%; Score 29; DB 2; Length 49;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 15 DDDDK 19

RESULT 6
T20731
hypothetical protein F11A1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20731
R:Harris, B.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z19316
A:Accession: T20731
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-92 <WIL>
A:Cross-references: EMBL:Z50857; PIDN:CAA90717.1; GSPDB:GN00028; CESP:F11A1.1
A:Experimental source: clone F11A1
C:Genetics:
A:Gene: CESP:F11A1.1
A:Map position: X
A:Introns: 58/3

Query Match 100.0%; Score 29; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 48 DDDDK 52

RESULT 7
T21867
hypothetical protein F36G9.15 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T21867
R:Wall, M.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19481
A:Accession: T21867
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-109 <WIL>
A:Cross-references: EMBL:Z81533; PIDN:CAB04329.1; GSPDB:GN00023; CESP:F36G9.15
A:Experimental source: clone F36G9
C:Genetics:
A:Gene: CESP:F36G9.15
A:Map position: 5
A:Introns: 10/3; 39/3; 66/2; 86/2
C:Superfamily: Caenorhabditis elegans hypothetical protein Y57G7A.3

Query Match 100.0%; Score 29; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 97 DDDDK 101

RESULT 8
T32477
hypothetical protein F52H2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T32477
R:Chissoe, S.; Hawkins, J.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid F52H2.

A:Reference number: Z21175
A:Accession: T32477
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-140 <CHI>
A:Cross-references: EMBL:AF026214; PIDN:AAB71313.1; GSPDB:GN00028; CESP:F52H2.3
A:Experimental source: strain Bristol N2; clone F52H2
C:Genetics:
A:Gene: CESP:F52H2.3
A:Map position: X
A:Introns: 45/3; 88/2

Query Match 100.0%; Score 29; DB 2; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
Db 125 DDDDK 129

RESULT 9

T33710

hypothetical protein F58E2.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 08-Sep-2000

C:Accession: T33710

R:Goela, D.; Delehaunty, A.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid F58E2.

A:Reference number: Z21390

A:Accession: T33710

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-141 <GOE>

A:Cross-references: EMBL:AF100659; PIDN:AAC58965.1; GSPDB:GN00022; CESP:F58E2.2

A:Experimental source: strain Bristol N2; clone F58E2

C:Genetics:

A:Gene: CESP:F58E2.2

A:Map position: 4

A:Introns: 83/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F58E2.2

Query Match 100.0%; Score 29; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
Db 55 DDDDK 59

RESULT 10

T23275

hypothetical protein K03D3.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000

C:Accession: T23275

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19720

A:Accession: T23275

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-144 <WTL>

A:Cross-references: EMBL:Z82276; PIDN:CAB05239.1; GSPDB:GN00022; CESP:K03D3.2

A:Experimental source: clone K03D3

C:Genetics:

A:Gene: CESP:K03D3.2

A:Map position: 4

A:Introns: 85/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F58E2.2

Query Match 100.0%; Score 29; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
Db 57 DDDDK 61

RESULT 11

A36237

melanin-concentrating hormone precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change 16-Jul-1999

C:Accession: A36237; A37407; A37406

R:Thompson, R.C.; Watson, S.J.

DNA Cell Biol. 9, 637-645, 1990

A:Title: Nucleotide sequence and tissue-specific expression of the rat melanin concen

A:Reference number: A36237; MUID:91083836; PMID:2261081

A:Accession: A36237

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-165 <THO>

A:Cross-references: GB:M62641; NID:G205329; PIDN:AAA1581.1; PID:G205330

R:Nahon, J.L.; Presse, F.; Bittencourt, J.C.; Sawchenko, P.E.; Vale, W.

Endocrinology 125, 2056-2065, 1989

A:Title: The rat melanin-concentrating hormone messenger ribonucleic acid encodes mu1

A:Reference number: A37407; MUID:90005146; PMID:2477226

A:Accession: A37407

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-165 <NAH>

A:Cross-references: GB:M29712; NID:G205327; PIDN:AAA1580.1; PID:G205328

R:Vaughan, J.M.; Fischer, W.H.; Hoeger, C.; Rivier, J.; Vale, W.

Endocrinology 125, 1660-1665, 1989

A:Title: Characterization of melanin-concentrating hormone from rat hypothalamus.

A:Reference number: A37406; MUID:89338286; PMID:2759038

A:Accession: A37406

A:Status: preliminary

A:Molecule type: protein

A:Residues: 147-165 <VAU>

C:Superfamily: melanin-concentrating hormone

Query Match 100.0%; Score 29; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
Db 76 DDDDK 80

RESULT 12

T33056

hypothetical protein F56C3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000

C:Accession: T33056

R:Stonking, T.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid F56C3.

A:Reference number: Z21276

A:Accession: T33056

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-166 <STO>

A:Cross-references: EMBL:AF067214; PIDN:AAC17006.1; GSPDB:GN00028; CESP:F56C3.3

A:Experimental source: strain Bristol N2; clone F56C3

C:Genetics:

A:Gene: CESP:F56C3.3

A:Map position: X

A:Introns: 15/1; 41/3; 71/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F56C3.3

T37039

hypothetical protein SCJ12.22 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37039
R:Murphy, L.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21619
A:Accession: T37039
A:Status: preliminary; translated from GB/EMBL/DDBY
A:Molecule type: DNA
A:Residues: 1-213 <MUR>
A:Cross-references: EMBL:AL109989; PIDN:CAB53434.1; GSPDB:GN00070; SCOEDB:SCJ12.22
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ12.22

Query Match 100.0%; Score 29; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 33 DDDDK 37

RESULT 18

TRBOTR
Trypsin (EC 3.4.21.4) precursor - bovine
N:Contains: trypsinogen
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 28-Feb-1986 #text_change 18-Jul-1997
C:Accession: A09164; A0946; S08774
R:Mikes, O.; Holeysovsky, V.; Tomasek, V.; Sorm, F.
Biochem. Biophys. Res. Commun. 24, 346-352, 1966
A:Title: Covalent structure of bovine trypsinogen. The position of the remaining amides.
A:Reference number: A90164; MUID:67168848; PMID:5367094

A:Accession: A90164
A:Molecule type: protein
A:Residues: 1-57,'Q',59-67,'Q',69-150,'N',152-176,'N',178-229 <MIK>
R:Titani, B.S.
Philos. Trans. R. Soc. Lond. B257, 77-87, 1970

A:Reference number: A93755
A:Contents: annotation; revisions
R:Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 14, 1358-1366, 1975
A:Title: Amino acid sequence of dogfish trypsin.
A:Reference number: A00950; MUID:75146445; PMID:1092332
A:Contents: annotation; revisions
A:Note: the sequence agrees with that shown
R:Bode, W.; Schwager, P.
J. Mol. Biol. 98, 693-717, 1975

A:Title: The refined crystal structure of bovine beta-trypsin at 1.8 angstrom resolution
A:Reference number: A92954; MUID:76072097; PMID:512
A:Contents: annotation; X-ray crystallography; binding sites for calcium, substrate, and
C:Comment: trypsinogen is synthesized in the acinar cells of the pancreas.
C:Comment: Autocatalytic cleavage after Lys-6 leads to beta-trypsin by releasing a terminal
s pseudotrypsin. A cleavage may also occur after Arg-105.
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-229/Product: trypsinogen #status experimental <ZYM>
F:1-6/Domain: activation peptide #status experimental <APT>
F:7-131,132-229/Product: trypsin homology <TRY>
F:6-7/Cleavage site: Lys-11e (enteropeptidase) #status experimental
F:13-143 31-47,115-216,122-189,154-168,179-203/Disulfide bonds: #status experimental
F:16,90,183/Active site: His, Asp, Ser #status experimental
F:58,60,63,68/Binding site: calcium (Glu, Asn, Val, Glu) #status experimental
F:131-132/Cleavage site: Lys-Ser (autolytic) #status experimental

Query Match 100.0%; Score 29; DB 1; Length 229;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 2 DDDDK 6

RESULT 19

TRDFS
Trypsin (EC 3.4.21.4) precursor - spiny dogfish
N:Alternate names: trypsinogen
C:Species: Squalus acanthias (spiny dogfish)
C:Date: 31-May-1979 #sequence_revision 31-May-1979 #text_change 21-Jan-1997
C:Accession: A00950; B27719
R:Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 14, 1358-1366, 1975
A:Title: Amino acid sequence of dogfish trypsin.
A:Reference number: A00950; MUID:75146445; PMID:1092332
A:Accession: A00950
A:Molecule type: protein
A:Residues: 8-229 <TRI>
A:Note: 119-Pro was also found
R:Hermodson, M.A.; Tye, R.W.; Reeck, G.R.; Neurath, H.; Walsh, K.A.
FEBS Lett. 14, 222-224, 1971
A:Title: Comparison of the amino terminal sequences of bovine, dogfish, and lungfish
A:Reference number: A27719
A:Accession: B27719
A:Molecule type: protein
A:Residues: 1-21 <HER>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase;
F:1-7/Domain: activation peptide #status experimental <APT>
F:8-229/Product: trypsin #status predicted <MAT>
F:8-223/Domain: trypsin homology <TRY>
F:14-143 32-48,116-216,123-189,154-168,179-203/Disulfide bonds: #status predicted
F:47,91,183/Active site: His, Asp, Ser #status predicted
F:59,61,64,69/Binding site: calcium (Glu, Asp, Ala, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 229;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 3 DDDDK 7

RESULT 20

TRFCTR
Trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)
N:Contains: trypsinogen
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 31-Mar-2000
C:Accession: A90641; A90368; A00947
R:Charles, M.; Roversy, M.; Guidoni, A.; Desnuelle, P.
Biochim. Biophys. Acta 69, 115-129, 1963
A:Title: Su le trypsinogene et la trypsine de porc.
A:Reference number: A90641
A:Accession: A90641
A:Molecule type: protein
A:Residues: 1-10 <CHA>
R:Hermodson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 12, 3146-3153, 1973
A:Title: Determination of the amino acid sequence of porcine trypsin by sequenator an
A:Reference number: A90368; MUID:73258692; PMID:4738933
A:Accession: A90368
A:Molecule type: protein
A:Residues: 9-231 <HER>
A:Note: at position 20, Ile and Val occur alternatively
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase;
F:1-231/Product: trypsinogen #status experimental <ZYM>
F:1-8/Domain: activation peptide #status experimental <APT>
F:9-231/Product: trypsin #status experimental <MAT>

F:9-224/Domain: trypsin homology <TRY>
 F:15-145,33-49,117-218,124-191,156-170,181-205/disulfide bonds: #status predicted
 F:48,92,185/Active site: His, Asp, Ser #status predicted
 F:60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 231;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 4 DDDDK 8

RESULT 21

T22698

hypothetical protein F55B12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Feb-2000

C:Accession: T22698

R:Simms, M.

submitted to the EMBL Data Library, September 1996

A:Reference number: Z19602

A:Accession: T22698

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-232 <WIL>

A:Cross-references: EMBL:Z79757; PIDN:CA802124.1; GSPDB:GN00023; CESP:F55B12.1

A:Experimental source: clone F55B12

C:Genetics:

A:Gene: CESP:F55B12.1

A:Map position: 5

A:Introns: 37/2; 59/1; 126/3; 168/2; 197/3

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 29; DB 2; Length 232;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 165 DDDDK 169

RESULT 22

T51270

hypothetical protein T8M16_150 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

C:Accession: T51270

R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.;

submitted to the Protein Sequence Database, August 2000

A:Reference number: Z25346

A:Accession: T51270

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-239 <BEN>

A:Cross-references: EMBL:AL390921

A:Experimental source: cultivar Columbia; BAC clone T8M16

C:Genetics:

A:Map position: 3

A:Introns: 113/1; 154/3; 173/3

A:Note: T8M16_150

Query Match 100.0%; Score 29; DB 2; Length 239;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 73 DDDDK 77

RESULT 23

T8T2

trypsin (EC 3.4.21.4) II precursor - rat

N:Alternate names: trypsinogen II

C:Species: Rattus norvegicus (Norway rat)

C:Date: 05-Apr-1983 #sequence_revision 30-Sep-1987 #text_change 18-Jul-1997

C:Accession: A22657; A00949

R:Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.

J. Biol. Chem. 259, 14255-14264, 1984

A:Title: Structure of two related rat pancreatic trypsin genes.

A:Reference number: A22657; MUID:85054880; PMID:6094547

A:Accession: A22657

A:Molecule type: DNA

A:Residues: 1-246 <CRA>

R:MacDonald, R.J.; Stary, S.J.; Swift, G.H.

J. Biol. Chem. 257, 9724-9732, 1982

A:Title: Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences

A:Reference number: A00948; MUID:82265624; PMID:6896710

A:Accession: A00949

A:Molecule type: mRNA

A:Residues: 9-246 <MAC>

C:Comment: The trypsin II mRNA is present in much lower quantities than the trypsin I

C:Genetics:

A:Introns: 14/1; 67/2

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-23/Domain: activation peptide #status predicted <APT>

F:24-246/Product: trypsin II #status predicted <ENZ>

F:24-239/Domain: trypsin homology <TRY>

F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted

F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 246;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 19 DDDDK 23

RESULT 24

TRDGC

trypsin (EC 3.4.21.4) precursor, cationic - dog

N:Alternate names: cationic trypsinogen

C:Species: Canis lupus familiaris (dog)

C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 18-Jun-1999

C:Accession: B26273

R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.

Mol. Cell Biol. 5, 2669-2676, 1985

A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA se

A:Reference number: A26273; MUID:86284628; PMID:3841794

A:Accession: B26273

A:Molecule type: mRNA

A:Residues: 1-246 <PIN>

A:Cross-references: GB:M11590; NID:g164096; PIDN:AAA30900.1; PID:g164097

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-23/Domain: activation peptide #status predicted <APT>

F:24-246/Product: trypsin, cationic #status predicted <ENZ>

F:24-239/Domain: trypsin homology <TRY>

F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted

F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 246;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
|
Db 19 DDDDK 23

RESULT 25

trypsin (EC 3.4.21.4) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 22-Jun-1999
C:Accession: B25528
R:Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.
Nucleic Acids Res. 14, 8307-8330, 1986
A:Title: Sequence organisation and transcriptional regulation of the mouse elastase II a

A:Reference number: A93646; MUID:87066713; PMID:3641189
A:Accession: B25528
A:Molecule type: mRNA
A:Residues: 1-246 <STE>

A:Cross-references: GB:X04574; NID:g54918; PIDN:CAA28243.1; PID:g54919
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-246/Product: trypsin #status predicted <MAT>
F:24-239/Domain: trypsin homology <TRY>

F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
|
Db 19 DDDDK 23

RESULT 26

trypsin (EC 3.4.21.4) I precursor [validated] - human
N:Alternate names: trypsin, cationic; trypsinogen I
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
C:Accession: A23852; B61066; A43988
R:Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.
Gene 41, 305-310, 1986
A:Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human
A:Reference number: A91544; MUID:86221712; PMID:3011602
A:Accession: A23852
A:Molecule type: mRNA

A:Residues: 1-247 <EMI>
A:Cross-references: GB:M22612; NID:g521215; PIDN:AAA61231.1; PID:g521216
R:Kimland, M.; Russick, C.; Marks, W.H.; Borgstroem, A.
Clin. Chim. Acta 184, 31-46, 1989

A:Title: Immunoreactive anionic and cationic trypsin in human serum.

A:Reference number: A61066; MUID:90091010; PMID:2598466

A:Accession: B61066
A:Molecule type: protein

A:Residues: 16-43 <KIM>
R:Koivunen, E.; Huhtala, M.L.; Stenman, U.H.
J. Biol. Chem. 264, 14095-14099, 1989

A:Title: Human ovarian tumor-associated trypsin. Its purification and characterization
A:Reference number: A43988; MUID:89340515; PMID:2503510
A:Accession: A43988

A:Molecule type: protein

A:Residues: 16-54 <KOI>
A:Experimental source: mucinous ovarian tumor cyst fluid

C:Genetics:

A:Gene: GDB:PRSS1; TRY1
A:Cross-references: GDB:119620; OMIM:276000

A:Map position: 7q35-7q35

A:Note: The human genome contains at least ten trypsin genes or pseudogenes, at least two
C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; phosphoprotein; protein digestion; serine proteinase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-246/Product: trypsinogen I #status experimental <ZYM>
F:16-23/Domain: activation peptide #status experimental <APT>
F:24-246/Product: trypsin I #status predicted <ENZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,139-206,171-185,196-220/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
|
Db 19 DDDDK 23

RESULT 27

trypsin (EC 3.4.21.4) II precursor [validated] - human
N:Alternate names: trypsin 2; trypsin, anionic; trypsinogen II
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
C:Accession: B25852; A61066; B43988
R:Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K.
Gene 41, 305-310, 1986
A:Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding human
A:Reference number: A91544; MUID:86221712; PMID:3011602
A:Accession: B25852
A:Molecule type: mRNA

A:Residues: 1-247 <EMI>
A:Cross-references: GB:M27602; NID:g521217; PIDN:AAA61232.1; PID:g521218
R:Kimland, M.; Russick, C.; Marks, W.H.; Borgstroem, A.
Clin. Chim. Acta 184, 31-46, 1989
A:Title: Immunoreactive anionic and cationic trypsin in human serum.

A:Reference number: A61066; MUID:90091010; PMID:2598466
A:Accession: A61066

A:Molecule type: protein
A:Residues: 16-39, 'X', 41-42, 'XXXX', 47-49 <KIM>

R:Koivunen, E.; Huhtala, M.L.; Stenman, U.H.
J. Biol. Chem. 264, 14095-14099, 1989

A:Title: Human ovarian tumor-associated trypsin. Its purification and characterization
A:Reference number: A43988; MUID:89340515; PMID:2503510
A:Accession: B43988

A:Molecule type: protein
A:Residues: 16-49 <KOI>

A:Experimental source: mucinous ovarian tumor cyst fluid
C:Genetics:

A:Gene: GDB:PRSS2; TRY2
A:Cross-references: GDB:335289; OMIM:601564

A:Map position: 7q35-7q35

A:Note: the human genome contains at least ten trypsin genes or pseudogenes, at least
C:Superfamily: trypsin; trypsin homology

F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-246/Product: trypsinogen II #status experimental <ZYM>

F:16-23/Domain: activation peptide #status experimental <APT>
F:24-246/Product: trypsin II #status predicted <ENZ>

F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,171-185,196-220/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|
|
|
|
|
Db 19 DDDDK 23

RESULT 28

TRDG
trypsin (EC 3.4.21.4) precursor, anionic - dog
N:Alternate names: cationic trypsinogen
C:Species: Canis lupus familiaris (dog)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 18-Jun-1999
C:Accession: A26273
R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.
Mol. Cell. Biol. 5, 2669-2676, 1985
A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA sequence
A:Reference number: A26273; MUID:86284628; PMID:3841794
A:Accession: A26273
A:Molecule type: mRNA
A:Residues: 1-247 <PIN>
A:Cross-references: GB:M1589; NID:g164094; PIDN:AAA30899.1; PID:g164095
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-247/Product: trypsin, anionic #status predicted <ENZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
DB 19 DDDDK 23

RESULT 29

A27547
trypsin (EC 3.4.21.4) precursor, cationic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 22-Jun-1999
C:Accession: A27547
R:Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.
Biochemistry 26, 3081-3086, 1987
A:Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.
A:Reference number: A27547; MUID:87271609; PMID:3607011
A:Accession: A27547
A:Molecule type: mRNA
A:Residues: 1-247 <FLE>
A:Cross-references: GB:M16624; NID:g206498; PIDN:AAA41985.1; PID:g206499
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F:25-240/Domain: trypsin homology <TRY>
F:31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
F:64,108,201/Active site: His, Asp, Ser #status predicted
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
DB 20 DDDDK 24

RESULT 30

S13813
trypsin (EC 3.4.21.4) - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S13813
R:le Huerou, I.; Wicker, C.; Guilloteau, P.; Toullec, R.; Puigserver, A.
Eur. J. Biochem. 193, 767-773, 1990

A:Title: Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anion
A:Reference number: S13813; MUID:91065383; PMID:1701147
A:Accession: S13813
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-247 <HUE>
A:Cross-references: EMBL:X54703; NID:g829; PIDN:CAA38513.1; PID:g830
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:24-239/Domain: trypsin homology <TRY>
F:63,107,200/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 29; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
DB 19 DDDDK 23

RESULT 31

S12764
trypsin (EC 3.4.21.4) III precursor - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-Jun-1999
C:Accession: S12764
R:Tani, T.; Kawashima, I.; Mita, K.; Takiguchi, Y.
Nucleic Acids Res. 18, 1631, 1990
A:Title: Nucleotide sequence of the human pancreatic trypsinogen III cDNA.
A:Reference number: S12764; MUID:90221895; PMID:2326201
A:Accession: S12764
A:Molecule type: mRNA
A:Residues: 1-247 <TAN>
A:Cross-references: EMBL:X15505; NID:g37459; PIDN:CAA33527.1; PID:g37460
C:Genetics:

A:Gene: GDB:PRSS3; TRY3
A:Cross-references: GDB:335297
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; pancreas; protein digestion; serine proteinase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-21/Domain: activation peptide #status predicted <APT>
F:22-247/Product: trypsin III #status predicted <MAT>
F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 100.0%; Score 29; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
DB 19 DDDDK 23

RESULT 32

E84708
probable signal peptidase I [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84708

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84708
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-250 <STO>
A;Cross-references: GB:AE002093; NID:g1946373; PIDN:AAB63091.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g30440
A;Map position: 2

Query Match 100.0%; Score 29; DB 2; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 42 DDDDK 46

RESULT 33

T45613
SNAP25-like protein - Arabidopsis thaliana
N;Alternate names: protein F13G24.80
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45613
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, December 1999
A;Reference number: 223009
A;Accession: T45613
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-251 <BEV>
A;Cross-references: EMBL:AL133421
A;Experimental source: cultivar Columbia; BAC clone F13G24
C;Genetics:
A;Map position: 5
A;Introns: 116/3; 148/1; 188/3; 219/2
A;Note: F13G24.80

Query Match 100.0%; Score 29; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 29 DDDDK 33

RESULT 34

B29653
aspartic acid-rich protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 01-Dec-2000
C;Accession: B29653; T18401
R;Lenstra, R.; d'Auriol, L.; Andrieu, B.; Le Bras, J.; Galibert, F.
Biochem. Biophys. Res. Commun. 146, 368-377, 1987
A;Title: Cloning and sequencing of Plasmodium falciparum DNA fragments containing repeti
A;Reference number: A90134; MUID:87270765; PMID:3038111
A;Accession: B29653
A;Molecule type: DNA
A;Residues: 1-253 <LEN>
A;Cross-references: GB:M17028; NID:g160339; PID:g160342
R;Luersen, K.; Walter, R.D.; Muller, S.
Mol. Biochem. Parasitol. 96, 131-142, 1999
A;Title: The putative gamma-glutamylcysteine synthetase from Plasmodium falciparum conta
A;Reference number: Z18928; MUID:99151773; PMID:10029315
A;Accession: T18401
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-125, '1', 127-253 <LUE>
A;Cross-references: EMBL:AJ700696; NID:e1301264; PID:e1311496; PIDN:CAA07355.1

Query Match 100.0%; Score 29; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 232 DDDDK 236

RESULT 35

I38363
trypsin (EC 3.4.21.4) IV form b precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 23-Aug-1997
C;Accession: I38363; S37538
R;Wiegand, U.; Corbach, S.; Minn, A.; Kang, J.; Muller-Hill, B.
Gene 136, 167-175, 1993
A;Title: Cloning of the cDNA encoding human brain trypsinogen and characterization of
A;Reference number: I38363; MUID:94123994; PMID:8294000
A;Accession: I38363
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-259 <RES>
A;Cross-references: EMBL:X71345; NID:g405755
C;Genetics:
A;Introns: 25/3; 78/1; 162/3; 208/2
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;37-251/Domain: trypsin homology <TRY>
F;75,119,212/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 29; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 32 DDDDK 36

RESULT 36

A86272
protein F16A14.10 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A86272
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86272
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <STO>
A;Cross-references: GB:AE005172; NID:g8778388; PIDN:AAF79396.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16A14.10
A;Map position: 1

Query Match 100.0%; Score 29; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 43 DDDDK 47

RESULT 37

S64615

hypothetical protein YGR280c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G9534

C:Species: Saccharomyces cerevisiae
C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
C:Accession: S64615

R:Voet, M.; Voickaert, G.
submitted to the Protein Sequence Database, May 1996

A:Reference number: S64611

A:Accession: S64615

A:Molecule type: DNA

A:Residues: 1-271 <VO>

A:Cross-references: EMBL:Z73065; NID:gl323510; PID:e243251; PID:gl323511; GSPDB:GN000077;
A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YGR280c

A:Cross-references: SGD:S0003512

A:Map position: 7R

Query Match 100.0%; Score 29; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 163 DDDDK 167

RESULT 38

A86384

unknown protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A86384

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 815-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A86384

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-288 <STO>

A:Cross-references: GB:AE005172; NID:gl0092327; PIDN:AGI2738.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 100.0%; Score 29; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 130 DDDDK 134

RESULT 39

B40722

homeotic protein lin-32 - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Nov-2000

C:Accession: B40722; A40720

R:Wang, B.B.; Mueller-Immergluck, M.M.; Austin, J.; Robinson, N.T.; Chisholm, A.; Kenyon

Cell 74, 29-42, 1993

A:Title: A homeotic gene cluster patterns the anteroposterior body axis of Caenorhabditis

A:Reference number: A40722; MUID:93327429; PMID:8101474

A:Accession: B40722

A>Status: preliminary

A:Map position: V

A:Molecule type: mRNA

A:Residues: 1-288 <WAN>

A:Cross-references: GB:LI9248

R:Clark, S.G.; Chisholm, A.D.; Horvitz, H.R.

Cell 74, 43-55, 1993

A:Title: Control of cell fates in the central body region of Caenorhabditis elegans b

A:Reference number: A40720; MUID:93327430; PMID:8101475

A:Accession: A40720

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 36-288 <CLA>

A:Cross-references: GB:LI9639; NID:g304331; PIDN:AAC37168.1; PID:g304332

C:Genetics:

A:Gene: lin-32

C:Superfamily: unassigned homeobox proteins; homeobox homolog

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:201-257/Domain: homeobox homolog <HOX>

Query Match 100.0%; Score 29; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 157 DDDDK 161

RESULT 40

S64308

hypothetical protein YGR017w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein G4023

C:Species: Saccharomyces cerevisiae

C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002

C:Accession: S64308

R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64071

A:Accession: S64308

A:Molecule type: DNA

A:Residues: 1-297 <RTE>

A:Cross-references: EMBL:Z72802; NID:gl322983; PID:e243923; PID:gl322984; GSPDB:GN000

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YGR017w

A:Cross-references: SGD:S0003249

A:Map position: 7R

Query Match 100.0%; Score 29; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5

DB 156 DDDDK 160

RESULT 41

T37251

homeobox protein ceh-24, NK-2 class - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000

C:Accession: T37251

R:Harfe, B.; Fire, A.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21646

A:Accession: T37251

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-298 <HAR>

A:Cross-references: EMBL:AF026056; PIDN:AAB81844.1

C:Genetics:

A:Gene: ceh-24

A:Map position: V

C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 29; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0;

QY 1 DDDDK 5
|||||
Db 232 DDDDK 236

RESULT 42

S33496 trypsin (EC 3.4.21.4) IV form a - human

C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Aug-1995 #text_change 15-Aug-1997
C:Accession: S33496

R:Wiegand, U.; Corbach, S.; Minn, A.; Kang, J.; Mueller-Hill, B.
submitted to the EMBL Data Library, March 1993

A:Description: Identification, cloning and characterization of a cDNA encoding a human

A:Reference number: S33496

A:Accession: S33496

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-304 <WIE>

A:Cross-references: EMBL:X72781

C:Genetics:

A:Gene: GDB:PRSS4; TRY4

A:Cross-references: GDB:335300

A:Map position: 7q35-7q35

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:81-296/Domain: trypsin homology <TRY>

F:120,164,257/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 29; DB 2; Length 304;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 76 DDDDK 80

RESULT 43

T37437

Probable 34.9K protein - vaccinia virus (strain Ankara)

C:Species: vaccinia virus

A:Variety: strain Ankara

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 20-Jun-2000

C:Accession: T37437

R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.

submitted to the EMBL Data Library, March 1997

A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain

A:Reference number: Z20877

A:Accession: T37437

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-310 <ANT>

A:Cross-references: EMBL:U94848; PIDN:AAB96542.1

A:Experimental source: strain Ankara

C:Genetics:

A:Note: MVA164R

C:Superfamily: vaccinia virus probable 34.9K protein

Query Match 100.0%; Score 29; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 85 DDDDK 89

RESULT 44

T31799

hypothetical protein F29G9.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T31799

R:Langston, Y.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F29G9.

A:Reference number: Z21088

A:Accession: T31799

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-310 <LAN>

A:Cross-references: EMBL:AF016440; PIDN:AAB65905.1; GSPDB:GN00023; CESP:F29G9.4

A:Experimental source: strain Bristol N2; clone F29G9

C:Genetics:

A:Gene: CESP:F29G9.4

A:Map position: 5

A:Introns: 36/3; 203/3; 271/1

Query Match 100.0%; Score 29; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 4 DDDDK 8

RESULT 45

T34409

hypothetical protein F07E5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34409

R:Du, Z.; Goela, D.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid F07E5.

A:Reference number: Z21520

A:Accession: T34409

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-315 <DUZ>

A:Cross-references: EMBL:U80837; PIDN:AAB37902.1; GSPDB:GN00020; CESP:F07E5.2

A:Experimental source: strain Bristol N2; clone F07E5

C:Genetics:

A:Gene: CESP:F07E5.2

A:Map position: 2

A:Introns: 28/2

Query Match 100.0%; Score 29; DB 2; Length 315;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 308 DDDDK 312

Search completed: October 14, 2003, 13:12:56

Job time : 13.8182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:20 ; Search time 6.36364 seconds
(without alignments)
36.950 Million cell updates/sec

Title: US-09-856-050-19_COPY_36_40
Perfect score: 29
Sequence: 1 DDDDK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	29	100.0	1 TRYP_FELCA	P81071 felis silve
2	29	100.0	47 1 SCRA_PANIM	P55927 pandinus im
3	29	100.0	71 1 YMA6_CAEEL	P34451 caenorhabdi
4	29	100.0	165 1 MLGH_RAT	P14200 rattus norv
5	29	100.0	173 1 GNTH_COTJA	Q9dgd4 c gnodotro
6	29	100.0	177 1 ARG1_YEAST	P07249 saccharomyc
7	29	100.0	229 1 TRYP_SQUAC	P00764 squalus aca
8	29	100.0	231 1 TRYP_PIG	P00761 sus scrofa
9	29	100.0	243 1 TRY1_BOVIN	P00760 bos taurus
10	29	100.0	246 1 TRY1_CANFA	P06871 canis famil
11	29	100.0	246 1 TRY2_MOUSE	P07146 mus musculu
12	29	100.0	246 1 TRY2_RAT	P00763 rattus norv
13	29	100.0	247 1 TRY1_HUMAN	P07477 homo sapien
14	29	100.0	247 1 TRY2_BOVIN	Q29463 bos taurus
15	29	100.0	247 1 TRY2_CANFA	P06872 canis famil
16	29	100.0	247 1 TRY2_HUMAN	P07478 homo sapien
17	29	100.0	247 1 TRY3_RAT	P08426 rattus norv
18	29	100.0	251 1 SN29_ARATH	Q9s496 arabidopsis
19	29	100.0	253 1 ASP_PLAFLS	P13825 plasmodium
20	29	100.0	253 1 LI39_CAEEL	P34684 caenorhabdi
21	29	100.0	263 1 SN30_ARATH	Q9lmg8 arabidopsis
22	29	100.0	271 1 YG5W_YEAST	P53335 saccharomyc
23	29	100.0	279 1 SPBP_RAT	P08723 rattus norv
24	29	100.0	297 1 YG1B_YEAST	P53210 saccharomyc
25	29	100.0	304 1 TRY3_HUMAN	P35030 homo sapien
26	29	100.0	334 1 VA51_VACCC	P21069 vaccinia vi
27	29	100.0	334 1 VA51_VACCV	Q01219 vaccinia vi
28	29	100.0	334 1 VA51_VARV	P33858 variola vir
29	29	100.0	355 1 CWG2_SCHPO	P32434 schizosacch
30	29	100.0	381 1 YB2_YEAST	P38199 saccharomyc
31	29	100.0	406 1 METP_BIFLO	Q8g3h4 bifidobacte
32	29	100.0	451 1 QSEC_HAEIN	P45336 haemophilus
33	29	100.0	468 1 MK11_YEAST	P20484 saccharomyc

ALIGNMENTS

RESULT 1

ID	TRYP_FELCA	STANDARD;	PRT;	16 AA.
AC	P81071;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Trypsin precursor (EC 3.4.21.4) (Fragment).			
OS	Felis silvestris catus (Cat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.			
OX	NCBI_TaxID=9685;			
RN	[1]			
RP	SEQUENCE.			
RC	TISSUE=Pancreas;			
RX	MEDLINE=97235546; PubMed=9080665;			
RA	Steiner J.M., Medinger T.L., Williams D.A.;			
RT	"Purification and partial characterization of feline trypsin.";			
RL	Comp. Biochem. Physiol. 116B:87-93(1997).			
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
DR	InterPro: IPR001254; Ser-protease_Try.			
DR	PROSITE: PS00240; TRYPSIN_DOM; PARTIAL.			
DR	PROSITE: PS00134; TRYPSIN_HIS; PARTIAL.			
DR	PROSITE: PS00135; TRYPSIN_SER; PARTIAL.			
KW	Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.			
FT	PROPEP 1 8 ACTIVATION PEPTIDE.			
FT	CHAIN 9 >16 TRYPSIN.			
FT	NON_TER 16 16			
SQ	SEQUENCE 16 AA; 1825 MW; A6D751B58760A86 CRC64;			

Query Match 100.0%; Score 29; DB 1; Length 16;

Best Local Similarity 100.0%; Pred No. 4.4;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 DDDDK 5
Db	4 DDDDK 8

RESULT 2

ID	SCRA_PANIM	STANDARD;	PRT;	47 AA.
AC	P55927;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Potassium channel blocking toxin PITX-K-alpha precursor (Pandinotoxin alpha) (Potassium channel blocking toxin 2) (Pi2) (Fragment).			
GN	PITX-1.			
OS	Pandinus imperator (Emperor scorpion).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;			
OC	Scorpionoidea; Scorpionidae; Pandinus.			
OX	NCBI_TaxID=55084;			
RN	[1]			

34	29	100.0	497	1	VE2_HPV20	P50766 human papil
35	29	100.0	502	1	VE2_HPV25	P36787 human papil
36	29	100.0	528	1	CTK1_YEAST	Q03957 saccharomyc
37	29	100.0	585	1	TAI2_HUMAN	Q8WVQ3 homo sapien
38	29	100.0	585	1	TAI2_MOUSE	P59055 mus musculu
39	29	100.0	634	1	FTSH_AQUAE	O67077 aquifex aeo
40	29	100.0	650	1	PDR4_YEAST	P19880 saccharomyc
41	29	100.0	652	1	DYIN_DICDI	P54703 dictyosteli
42	29	100.0	658	1	GR78_XENLA	Q91883 xenopus lae
43	29	100.0	685	1	MDL1_CANAL	P97998 candida alb
44	29	100.0	786	1	FL10_CHLRE	P46869 chlamydomon
45	29	100.0	790	1	RA16_YEAST	P31244 saccharomyc

```
RP SEQUENCE FROM N.A.
RA Rogowski R.S., Collins J.H., O'Neill T.J., Gustafson T.A.,
RA Werkman T.R., Rogawski M.A., Tenenholz T.C., Weber D.J.,
RA Blaustein M.P.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 13-47.
RC TISSUE=Venom;
RX MEDLINE=97070422; PubMed=8913348;
RA Rogowski R.S., Collins J.H., O'Neill T.J., Gustafson T.A.,
RA Werkman T.A., Rogawski M.A., Tenenholz T.C., Weber D.J.,
RA Blaustein M.P.;
RT "Three new toxins from the scorpion Pandinus imperator selectively
RT block certain voltage-gated K+ channels.";
RL Mol. Pharmacol. 50:1167-1177(1996).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=97215813; PubMed=9062103;
RA Tenenholz T.C., Rogowski R.S., Collins J.H., Blaustein M.P.,
RA Weber D.J.;
RT "Solution structure for Pandinus toxin K-alpha (PITX-K alpha), a
RT selective blocker of A-type potassium channels.";
RL Biochemistry 36:2763-2771(1997).
CC -1- FUNCTION: Potent inhibitor of the A-type voltage-gated potassium
CC channels. Most potent inhibitor of Kv1.2 channels.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: BELONGS TO THE SHORT SCORPION TOXIN FAMILY. POTASSIUM
CC CHANNEL INHIBITORS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U79579; AB52576.1; -.
DR PIR: T10471; T10471.
DR PDB: 2PTA; 10-DEC-97.
DR InterPro: IPR001947; Scorpion_toxins.
DR Pfam: PF00451; toxin_2; 1.
DR ProDom: PD003586; Scorpion_toxins; 1.
DR PROSITE: PS01138; SCORP_SHORT_TOXIN; 1.
KW Toxin; Neurotoxin; Ionic channel inhibitor;
KW Potassium channel inhibitor; 3D-structure.
FT NON_TER 1 1
FT PEPTIDE 13 47 POTASSIUM CHANNEL BLOCKING TOXIN PITX-K-
FT ALPHA.
FT DISULFID 16 37
FT DISULFID 22 42
FT DISULFID 26 44
FT HELIX 19 29
FT TURN 30 30
FT STRAND 35 38
FT STRAND 39 40
FT STRAND 41 44
SQ SEQUENCE 47 AA; 5434 MW; 21B8BFF10A37BD7C CRC64;

Query Match 100.0%; Score 29; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 8 DDDDK 12

RESULT 3
YMA6_CAEEL STANDARD; PRT; 71 AA.
AC P34451;

SEQUENCE FROM N.A.
RA Rogowski R.S., Collins J.H., O'Neill T.J., Gustafson T.A.,
RA Werkman T.R., Rogawski M.A., Tenenholz T.C., Weber D.J.,
RA Blaustein M.P.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 13-47.
RC TISSUE=Venom;
RX MEDLINE=97070422; PubMed=8913348;
RA Rogowski R.S., Collins J.H., O'Neill T.J., Gustafson T.A.,
RA Werkman T.A., Rogawski M.A., Tenenholz T.C., Weber D.J.,
RA Blaustein M.P.;
RT "Three new toxins from the scorpion Pandinus imperator selectively
RT block certain voltage-gated K+ channels.";
RL Mol. Pharmacol. 50:1167-1177(1996).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=97215813; PubMed=9062103;
RA Tenenholz T.C., Rogowski R.S., Collins J.H., Blaustein M.P.,
RA Weber D.J.;
RT "Solution structure for Pandinus toxin K-alpha (PITX-K alpha), a
RT selective blocker of A-type potassium channels.";
RL Biochemistry 36:2763-2771(1997).
CC -1- FUNCTION: Potent inhibitor of the A-type voltage-gated potassium
CC channels. Most potent inhibitor of Kv1.2 channels.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: BELONGS TO THE SHORT SCORPION TOXIN FAMILY. POTASSIUM
CC CHANNEL INHIBITORS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23645; AAK26136.1; -.
DR WormPep; F54F2.6; CE00199.
KW Hypothetical protein.
SQ SEQUENCE 71 AA; 8546 MW; 15A7BDE6DE8C2DF8 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 34 DDDDK 38

RESULT 4
MLCH_RAT
ID MLCH_RAT STANDARD; PRT; 165 AA.
AC P14200;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pro-MCH precursor [Contains: Neuropeptide-glycine-glutamic acid (NGE)
DE (Neuropeptide G-E); Neuropeptide-glutamic acid-isoleucine (NEI)
DE (Neuropeptide E-I); Melanin-concentrating hormone (MCH)].
GN PMCH OR MCH.
OS Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RX MEDLINE=90005146; PubMed=2477226;
RA Nahon J.-L., Presse F., Bittencourt J.C., Sawchenko P.E., Vale W.;
RT "The rat melanin-concentrating hormone messenger ribonucleic acid
RT encodes multiple putative neuropeptides coexpressed in the
RT dorsolateral hypothalamus.";
```


RL Endocrinology 125:2056-2065(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91083836; PubMed-2261081;
 RA Thompson R.C., Watson S.J.;
 RT "Nucleotide sequence and tissue-specific expression of the rat
 RL melanin concentrating hormone gene.";
 RN DNA Cell Biol. 9:637-645(1990).
 RP [3]
 RP SEQUENCE OF 147-165.
 RC TISSUE-HYPOTHALAMUS;
 RX MEDLINE-89338286; PubMed-2759038;
 RA Vaughan J.M., Fischer W.H., Hoeger C., Rivier J., Vale W.;
 RT "Characterization of melanin-concentrating hormone from rat
 RL hypothalamus.";
 RN Endocrinology 125:1660-1665(1989).
 RP [4]
 RP TISSUE SPECIFICITY.
 RC STRAIN-Sprague-Dawley, and Wistar;
 RX MEDLINE-95303246; PubMed-7783849;
 RA Hervieu G., Nahon J.-L.;
 RT "Pro-melanin concentrating hormone messenger ribonucleic acid and
 RL peptides expression in peripheral tissues of the rat.";
 RN Neuroendocrinology 61:348-364(1995).
 RP [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE-96146469; PubMed-8593803;
 RA Hervieu G., Volant K., Grishina O., Descroix-Vagne M., Nahon J.-L.;
 RT "Similarities in cellular expression and functions of melanin-
 RL concentrating hormone and atrial natriuretic factor in the rat
 RT digestive tract.";
 RN Endocrinology 137:561-571(1996).
 RP [6]
 RP TISSUE SPECIFICITY.
 RX MEDLINE-95342335; PubMed-7617126;
 RA Takahashi K., Suzuki H., Totsumi K., Murakami O., Satoh F., Sone M.,
 RA Sasano H., Mouru T., Shibahara S.;
 RT "Melanin-concentrating hormone in human and rat.";
 RN Neuroendocrinology 161:493-498(1995).
 RP [7]
 RP PROCESSING, AND AMIDATION OF NEI.
 RX MEDLINE-93010665; PubMed-1327720;
 RA Parkes D., Vale W.;
 RT "Secretion of melanin-concentrating hormone and neuropeptide-BI from
 RL cultured rat hypothalamic cells.";
 RN Endocrinology 131:1826-1831(1992).
 RP [8]
 RP PROCESSING.
 RX MEDLINE-99156937; PubMed-10037747;
 RA Viale A., Ortolia C., Hervieu G., Furuta M., Barbero P., Steiner D.F.,
 RA Seidah N.G., Nahon J.-L.;
 RT "Cellular localization and role of prohormone convertases in the
 RL processing of pro-melanin concentrating hormone in mammals.";
 RN J. Biol. Chem. 274:6536-6545(1999).
 RP [9]
 RP FUNCTION OF NEI.
 RX MEDLINE-95375766; PubMed-7647772;
 RA Bluet-Pajot M.T., Presse F., Voko Z., Hoeger C., Mounier F.,
 RA Epelbaum J., Nahon J.-L.;
 RT "Neuropeptide-BI antagonizes the action of melanin-concentrating
 RL hormone on stress-induced release of adrenocorticotropin in the rat.";
 RN J. Neuroendocrinol. 7:297-303(1995).
 RP [10]
 RP FUNCTION OF NEI.
 RX MEDLINE-98366105; PubMed-9700748;
 RA Gonzalez M.I., Baker B.I., Hole D.R., Wilson C.A.;
 RT "Behavioral effects of neuropeptide E-I (NEI) in the female rat:
 RL interactions with alpha-MSH, MCH and dopamine.";
 RN Peptides 19:1007-1016(1998).
 RP [11]
 RP FUNCTION OF MCH AND NEI.
 RX MEDLINE-99025517; PubMed-9809645;
 RA Kistler-Heer V., Schlumpf M., Lichtensteiger W.;

RT "Melanocortin and MCH precursor-derived NEI effects on striatum-
 RL midbrain co-cultures.";
 RN Peptides 19:1317-1327(1998).
 CC -|- FUNCTION: MCH inhibits acth secretion at the end of the light on
 CC period which corresponds to the peak of the circadian rhythm in
 CC ACTH. Inhibits also stress induced ACTH release during the light
 CC off period of the cycle. Involved as a neurotransmitter or
 CC neuromodulator in a broad array of neuronal functions. Stimulates
 CC sexual behavior when injected into the ventromedial nucleus, this
 CC effect is antagonized by NEI. In the medial preoptic area,
 CC stimulates anxiety and sexual behavior. Antagonizes inhibitory
 CC effect of melanotropin alpha on exploration behavior.
 CC -|- FUNCTION: NEI CAN INFLUENCE DIFFERENTIATION OF NEURONAL PROCESSES
 CC IN BRAIN NEURONS. AFFECTS THE CONTENT OF NEUROFILAMENT PROTEIN IN
 CC NEURITOGENESIS (IN VITRO). MAY ALSO BE A NEUROMODULATORY FACTOR.
 CC IN BEHAVIORAL TESTS, IT STIMULATES EXPLORATION AND ANXIETY WHEN
 CC INJECTED INTO THE VENTROMEDIAL NUCLEUS. ALSO STIMULATES GROOMING,
 CC LOCOMOTION AND REARING. MAY ANTAGONIZE THE INHIBITORY EFFECT OF
 CC MCH ON ACTH RELEASE. REDUCES DOPAMINE AND DOPAC RELEASE IN THE
 CC VENTROMEDIAL NUCLEUS.
 CC -|- TISSUE SPECIFICITY: MCH IS PRESENT IN ALL REGIONS OF THE BRAIN AND
 CC IN NEUROINTERMEDIATE LOBE OF THE PITUITARY GLAND, WITH HIGHEST
 CC CONCENTRATIONS IN THE HYPOTHALAMUS. ALSO EXPRESSED TO A MUCH
 CC LESSER EXTENT IN STOMACH, LAMINA PROPRIA OF BOTH DUODENUM AND
 CC COLON, OVARY, THYMUS, PANCREAS, ADRENAL GLAND AND TESTIS
 CC (SPERMATOGONIA, EARLY SPERMATOCYTES AND SERTOLI CELLS). WEAK
 CC EXPRESSION IN HEART AND LUNG. THE OTHER PEPTIDES ARE EXPRESSED AT
 CC LEAST IN SERTOLI CELLS, NEI BEING ALSO EXPRESSED IN BRAIN, STOMACH
 CC AND PROXIMAL DUODENUM. IN BRAIN EXCLUSIVELY MATURE MCH AND NEI
 CC PEPTIDES ARE PRESENT. IN PERIPHERAL TISSUES A LARGE PRODUCT,
 CC ENCOMPASSING THE NEI AND MCH DOMAINS OF THE PRECURSOR, IS FOUND
 CC PREDOMINANTLY. AT LOW LEVELS FULLY PROCESSED MCH AND NEI PEPTIDES
 CC ARE PRESENT IN GUT. NO EXPRESSION IN PERIPHERAL BLOOD.
 CC -|- DEVELOPMENTAL STAGE: EXPRESSION IS STRONGLY INCREASED IN
 CC HYPOTHALAMUS BETWEEN POSTNATAL DAYS 12 AND 20, TO REACH HIGH
 CC CONSTANT VALUES IN ADULT.
 CC -|- INDUCTION: INHIBITED BY NEUROGENIC STRESS OR OSMOTIC STRESS.
 CC -|- PTM: PRO-MCH IS PROCESSED DIFFERENTIALLY IN THE BRAIN AND IN
 CC PERIPHERAL ORGANS PRODUCING TWO NEUROPEPTIDES: NEI AND MCH. A
 CC THIRD PEPTIDE, NGE, MAY ALSO BE PRODUCED. PREFERENTIAL PROCESSING
 CC IN NEURONS BY PROHORMONE CONVERTASE 2 (PC2) GENERATES NEI. MCH IS
 CC GENERATED IN NEURONS OF THE LATERAL HYPOTHALMIC AREA BY SEVERAL
 CC PROHORMONE CONVERTASES INCLUDING PC1/3, PC2 AND PC5/6.
 CC -|- PTM: MCH IS A CYCLIC PEPTIDE.
 CC -|- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M29712; AAA41580.1; -;
 CC EMBL; M62641; AAA41581.1; -;
 CC PIR; A36237; A36237.
 CC InterPro: IPR005456; Ppmelanin_conc.
 CC PRINTS; PR01641; PROMCHFAMILY.
 CC Cleavage on pair of basic residues; Hormone; Neuropeptide; Signal;
 KW Amidation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 165 PRO-MCH.
 FT PEPTIDE 110 128 NGE (POTENTIAL).
 FT PEPTIDE 131 143 NEI.
 FT PEPTIDE 147 165 MELANIN-CONCENTRATING HORMONE.
 FT MOD_RES 143 143 AMIDATION (G-144 PROVIDE AMIDE GROUP).
 FT DISULFID 153 162 BY SIMILARITY.
 SQ SEQUENCE 165 AA; 18482 MW; 38E9F07693E77A05 CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 165;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 DDDDK 5
DB      76 DDDDK 80

RESULT 5
GNIH_COTJA
ID      GNIH_COTJA      STANDARD;      PRT;      173 AA.
AC      Q9DGD4;
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DE      Gonadotropin inhibitory hormone precursor [Contains: Gonadotropin
DE      inhibitory hormone (GnIH); Gonadotropin inhibitory hormone-related
DE      peptide 1 (GnIH-RP1); Gonadotropin inhibitory hormone-related peptide
DE      2 (GnIH-RP2)].
GN      GNIH.
OS      Coturnix coturnix japonica (Japanese quail).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Coturnix.
OX      NCBI_TaxID=93934;
RN      [1]
RP      SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND MASS SPECTROMETRY.
RC      TISSUE=Brain;
RX      MEDLINE=21092948; PubMed=11171117;
RA      Satake H., Hisada M., Kawada T., Minakata H., Ukena K., Tsutsui K.;
RT      "Characterization of a cDNA encoding a novel avian hypothalamic
RT      neuropeptide exerting an inhibitory effect on gonadotropin release.";
RL      Biochem. J. 354:379-385(2001).
RN      [2]
RP      SEQUENCE OF 104-115, SYNTHESIS, FUNCTION, TISSUE SPECIFICITY, AND MASS
RP      SPECTROMETRY.
RC      TISSUE=Brain;
RX      MEDLINE=20422313; PubMed=10964719;
RA      Tsutsui K., Saigoh E., Ukena K., Teranishi H., Fujisawa Y.,
RA      Kikuchi M., Ishii S., Sharp P.J.;
RT      "A novel avian hypothalamic peptide inhibiting gonadotropin release.";
RL      Biochem. Biophys. Res. Commun. 275:661-667(2000).
CC      -!- FUNCTION: Hypothalamic factor, responsible for the negative
CC      regulation of gonadotropin secretion.
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -!- TISSUE SPECIFICITY: Specifically expressed in the dienecephalon.
CC      -!- MASS SPECTROMETRY: MW=1390.81; METHOD=WALDI; RANGE=104-115.
CC      -!- MASS SPECTROMETRY: MW=1389.4; METHOD=FAB; RANGE=104-115.
CC      -!- MASS SPECTROMETRY: MW=1501.81; METHOD=WALDI; RANGE=142-154.
CC      -!- SIMILARITY: BELONGS TO THE FARP (FMRP/AMIDE RELATED PEPTIDE)
CC      FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR      EMBL; AB039815; BABI5932.1; -.
DR      Neuropeptide; Signal; Amidation.
FT      SIGNAL      1      26      POTENTIAL.
FT      PROPEP      27      82
FT      PEPTIDE      84      95      GONADOTROPIN INHIBITORY HORMONE-RELATED
FT                                     PEPTIDE 1 (POTENTIAL).
FT
FT      PROPEP      98      103      GONADOTROPIN INHIBITORY HORMONE.
FT      PEPTIDE      104      115
FT      PROPEP      118      140      GONADOTROPIN INHIBITORY HORMONE-RELATED
FT      PEPTIDE      142      154      PEPTIDE 2.
FT
FT      PROPEP      157      173
FT      MOD_RES      95      95      (POTENTIAL).
FT      MOD_RES      115      115      AMIDATION (G-96 PROVIDE AMIDE GROUP)
FT
FT      MOD_RES      115      115      AMIDATION (G-116 PROVIDE AMIDE GROUP).
-----
FT      MOD_RES      154      154      AMIDATION (G-155 PROVIDE AMIDE GROUP)
FT      SEQUENCE      173 AA;      19428 MW;      B8D5CD3239C8E61B CRC64;
SQ      (POTENTIAL).
Query Match      100.0%;      Score 29;      DB 1;      Length 173;
Best Local Similarity      100.0%;      Pred. No. 52;
Matches      5;      Conservative      0;      Mismatches      0;      Indels      0;      Gaps      0;
QY      1 DDDDK 5
DB      40 DDDDK 44

RESULT 6
ARG1_YEAST
ID      ARG1_YEAST      STANDARD;      PRT;      177 AA.
AC      P07249;
DT      01-APR-1988 (Rel. 07, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Arginine metabolism regulation protein 1.
DE      Arginine metabolism regulation protein 1.
GN      ARG1 OR ARG80 OR YMR042W OR YW9532.07.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87257286; PubMed=3298999;
RA      Dubois E., Bercy J., Messenguy F.;
RT      "Characterization of two genes, ARG1 and ARGRIII required for
RT      specific regulation of arginine metabolism in yeast.";
RL      Mol. Gen. Genet. 207:142-148(1987).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / AB972;
RX      PubMed=9169872;
RA      Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA      Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA      Jagsels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA      Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT      "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      XIII.";
RL      Nature 387:90-93(1997).
CC      -!- FUNCTION: THE ARG1, ARG2 AND ARG3 REGULATORY PROTEINS ARE
CC      NECESSARY FOR THE REPRESSION OF SIX ANABOLIC GENES AND THE
CC      INDUCTION OF TWO CATABOLIC GENES BY ARGinine.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC      FACTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR      EMBL; X05327; CAA28944.1; -.
DR      EMBL; Z48502; CAA88408.1; -.
DR      PIN; S05822; RBYGI.
DR      HSP; P11746; IMNM.
DR      TRANSFAC; T00043; -.
DR      SGD; S0004645; ARG80.
DR      GO; GO:0005634; C:nucleus; IDA.
DR      GO; GO:0003704; F:Specific RNA polymerase II transcription fa...; IDA.
DR      GO; GO:0045944; P:Positive regulation of transcription from P...; IDA.
DR      InterPro; IPR002100; TF_MADSbox.
DR      Pfam; PF00319; SRF-TF; 1.
DR      PRINTS; PR00404; MADSDOMAIN.
DR      SMART; SM00432; MADS; 1.
DR      PROSITE; PS00350; MADS_BOX_1; 1.

```

DR PROSITE; PS50066; MADS_BOX_2; 1.
 KW Transcription regulation; DNA-binding; Arginine metabolism;
 FT Nuclear protein.
 FT DOMAIN 80 134 MADS.
 SQ SEQUENCE 177 AA; 19487 MW; 1F46C5C53ED8EE6 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 177;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 50 DDDDK 54

RESULT 7
 TRYP_SQUAC STANDARD; PRT; 229 AA.
 ID TRYP_SQUAC
 AC P00764;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trypsin precursor (EC 3.4.21.4).
 OS Squalus acanthias (Spiny dogfish).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Squaloidei; Squalidae; Squalus.
 OX NCBI_TaxID=7797;
 RN [1]
 RP SEQUENCE OF 8-229.
 RX MEDLINE=75146445; PubMed=1092332;
 RA Titani K., Ericsson L.H., Neurath H., Walsh K.A.;
 RT "Amino acid sequence of dogfish trypsin."
 RL Biochemistry 14:1358-1366(1975).
 RN [2]
 RP SEQUENCE OF 1-21.
 RA Hermodson M.A., Tye R.W., Reek G.R., Neurath H., Walsh K.A.;
 RT "Comparison of the amino terminal sequences of bovine, dogfish, and lungfish trypsins."
 RL FEBS Lett. 14:222-224(1971).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR PIR; A0950; TRDFS.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.151; .
 DR InterPro; IPR001254; Ser_protease_Try.
 DR PFAM; PF00089; trypsin; 1.
 DR SMART; SM0020; TRYP_SPC; 1.
 DR PROSITE; PS50240; TRYP_SIN_DOM; 1.
 DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
 DR PROSITE; PS00135; TRYP_SIN_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen.
 FT PROPEP 1 7 ACTIVATION PEPTIDE.
 FT CHAIN 8 229 TRYP_SIN.
 FT ACT_SITE 47 91 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 91 97 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 183 183 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 14 143 BY SIMILARITY.
 FT DISULFID 32 48 BY SIMILARITY.
 FT DISULFID 116 216 BY SIMILARITY.
 FT DISULFID 123 189 BY SIMILARITY.
 FT DISULFID 154 168 BY SIMILARITY.
 FT DISULFID 179 203 BY SIMILARITY.
 FT SITE 177 177 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
 FT VARIANT 109 109 L -> P.
 SQ SEQUENCE 229 AA; 24591 MW; E83B83C5AD72FCE4 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 229;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 50 DDDDK 54

Db 3 DDDDK 7
 RESULT 8
 TRYP_PIG STANDARD; PRT; 231 AA.
 ID TRYP_PIG
 AC P00761;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trypsin precursor (EC 3.4.21.4).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE OF 1-10.
 RA Charles M., Rovey M., Guidoni A.A., Desnuelle P.;
 RT "On trypsinogen and trypsin of pig."
 RL Biochim. Biophys. Acta 69:115-129(1963).
 RN [2]
 RP SEQUENCE OF 9-231.
 RX MEDLINE=73258692; PubMed=4738933;
 RA Hermodson M.A., Ericsson L.H., Neurath H., Walsh K.A.;
 RT "Determination of the amino acid sequence of porcine trypsin by sequenator analysis."
 RL Biochemistry 12:3146-3153(1973).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=93187998; PubMed=8445634;
 RA Huang Q., Liu S., Tang Y.;
 RT "Refined 1.6-A resolution crystal structure of the complex formed between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the squash family. Detailed comparison with bovine beta-trypsin and its complex."
 RL J. Mol. Biol. 229:1022-1030(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=92201369; PubMed=1551419;
 RA Huang Q., Liu S., Tang Y., Zeng F., Qian R.;
 RT "Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray crystal structure of its complex with porcine beta-trypsin."
 RL FEBS Lett. 297:143-146(1992).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=95035057; PubMed=7947985;
 RA Huang Q., Wang Z., Li Y., Liu S., Tang Y.;
 RT "Refined 1.8-A resolution crystal structure of the porcine epsilon-trypsin."
 RL Biochim. Biophys. Acta 1209:77-82(1994).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.
 RX MEDLINE=97390427; PubMed=9242660;
 RA Stubbs M.T., Morenweiser R., Sturzebecher J., Bauer M., Bode W., Huber R., Piechottka G.P., Matschner G., Sommerhoff C.P., Fritz H., Auerwald E.A.;
 RT "The three-dimensional structure of recombinant leech-derived trypsinase inhibitor in complex with trypsin. Implications for the structure of human mast cell trypsinase and its inhibition."
 RL J. Biol. Chem. 272:19931-19937(1997).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
 RX MEDLINE=98046095; PubMed=9384562;
 RA di Marco S., Priestle J.P.;
 RT "Structure of the complex of leech-derived trypsinase inhibitor (LDTI) with trypsin and modeling of the LDTI-trypsinase system."
 RL Structure 5:1465-1474(1997).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR PIR; A90641; TRPGTR.
 DR PDB; IMCT; 3I-JAN-94.
 DR PDB; 1AKS; 12-FEB-97.

DR PDB; 1EPT; 07-FEB-95.
 DR PDB; 1TFX; 21-JAN-98.
 DR PDB; 1LDT; 20-MAY-98.
 DR PDB; 1AN1; 01-JUL-98.
 DR PDB; 1AVW; 18-NOV-98.
 DR PDB; 1AVX; 18-NOV-98.
 DR PDB; 1C9P; 26-SEP-01.
 DR PDB; 1D30; 30-NOV-99.
 DR PDB; 1DF2; 26-SEP-01.
 DR PDB; 1EJA; 02-MAR-01.
 DR PDB; 1EWU; 28-JUN-00.
 DR PDB; 1FWG; 07-NOV-01.
 DR PDB; 1FN6; 07-NOV-01.
 DR PDB; 1FNI; 07-NOV-01.
 DR PDB; 1LT2; 05-JUN-02.
 DR PDB; 1LOU; 14-JUN-00.
 DR MEROPS; S01.151; --.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
 KW 3D-structure.

FT PROPEP 1 8
 FT CHAIN 9 231
 FT ACT_SITE 48 48
 FT ACT_SITE 92 92
 FT ACT_SITE 185 185
 FT DISULFID 15 145
 FT DISULFID 33 49
 FT DISULFID 117 218
 FT DISULFID 124 191
 FT DISULFID 156 170
 FT DISULFID 181 205
 FT SITE 179 179
 FT VARIANT 20 20
 FT STRAND 10 10
 FT STRAND 13 14
 FT TURN 17 18
 FT TURN 21 22
 FT TURN 23 27
 FT STRAND 31 39
 FT TURN 40 41
 FT STRAND 42 45
 FT STRAND 47 49
 FT STRAND 55 58
 FT STRAND 62 62
 FT TURN 63 64
 FT STRAND 71 80
 FT TURN 82 83
 FT TURN 86 88
 FT TURN 90 91
 FT STRAND 94 98
 FT STRAND 112 112
 FT TURN 120 121
 FT STRAND 123 128
 FT STRAND 142 142
 FT STRAND 144 150
 FT STRAND 153 159
 FT TURN 161 163
 FT TURN 166 166
 FT TURN 168 167
 FT STRAND 168 171
 FT TURN 174 175
 FT STRAND 179 179
 FT TURN 182 183
 FT TURN 185 186
 FT STRAND 188 191
 FT TURN 192 193
 FT STRAND 194 202
 FT STRAND 207 207
 FT TURN 208 209

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
 I -> V.

FT STRAND 210 210
 FT STRAND 212 216
 FT HELIX 217 219
 FT HELIX 221 231
 SQ SEQUENCE 231 AA; 24409 MW; A0A125CF7FC138C2 CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 231;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 DB 4 DDDDK 8
 RESULT 9
 TRY1_BOVIN
 ID TRY1_BOVIN STANDARD; PRT; 243 AA.
 AC P00760;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trypsinogen, cationic precursor (EC 3.4.21.4) (Beta-trypsin)
 DE (Fragment).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Okajima T., Maniwa M., Nagao S., Fujikawa H., Goto S.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 15-243, AND DISULFIDE BONDS.
 RX MEDLINE=67168848; PubMed=5967094;
 RA Mikes O., Holeysovsky V., Tomasek V., Sorm F.;
 RT "Covalent structure of bovine trypsinogen. The position of the
 RL remaining amides.";
 RL Biochem. Biophys. Res. Commun. 24:346-352(1966).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=72035053; PubMed=4399051;
 RA Hartley B.S.;
 RT "Homologies in serine proteinases.";
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:77-87(1970).
 RN [4]
 RP REVISIONS.
 RX MEDLINE=75146445; PubMed=1092332;
 RA Titani K., Ericsson L.H., Neurath H., Walsh K.A.;
 RT "Amino acid sequence of dogfish trypsin.";
 RL Biochemistry 14:1358-1366(1975).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF CALCIUM-BINDING SITE.
 RX MEDLINE=76072097; PubMed=512;
 RA Bode W., Schwager P.;
 RT "The refined crystal structure of bovine beta-trypsin at 1.8-A
 resolution. II. Crystallographic refinement, calcium binding site,
 RT benzamidine binding site and active site at pH 7.0.";
 RL J. Mol. Biol. 98:693-717(1975).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=77112431; PubMed=556951;
 RA Kossiakoff A.A., Chambers J.L., Kay L.M., Stroud R.M.;
 RT "Structure of bovine trypsinogen at 1.9-A resolution.";
 RL Biochemistry 16:654-664(1977).
 RN [7]
 RP DISULFIDE BONDS.
 RX MEDLINE=66079271; PubMed=5892911;
 RA Kauffman D.L.;
 RT "The disulphide bridges of trypsin.";
 RL J. Mol. Biol. 12:929-932(1965).


```

Best Local Similarity 100.0%; Pred. No. 74;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 16 DDDDK 20

RESULT 10
ID TRY1_CANFA STANDARD; PRT; 246 AA.
AC P06871;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin, cationic precursor (EC 3.4.21.4).
OS Canis familiaris (Dog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284628; PubMed=3841794;
RA Pinsky S.D., Laforge K.S., Scheele G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT in the dog pancreas."
RL Mol. Cell. Biol. 5:2669-2676(1985).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M11590; AAA30900.1; -
DR PIR: B26273; TRDGC.
DR HSP: P00761; 1EPT.
DR MEROPS: S01.151; -
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 23 ACTIVATION PEPTIDE.
FT CHAIN 24 246 TRYPSIN, CATIONIC.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 160 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 132 233 BY SIMILARITY.
FT DISULFID 139 206 BY SIMILARITY.
FT DISULFID 171 185 BY SIMILARITY.
FT DISULFID 196 220 BY SIMILARITY.
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 246 AA; 26170 MW; E9E5A1DE2391BBBB CRC64;

Query Match 100.0%; Score 29; DB 1; Length 246;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 16 DDDDK 20

us-09-856-050-19_copy_36_40.rsp
Page 8

```

RESULT 12
 ID TRY2_RAT TRY2_RAT STANDARD; PRT; 246 AA.
 AC P00763;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trypsin II, anionic precursor (EC 3.4.21.4) (Pretirpsinogen II).
 GN TRY2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85054880; PubMed=6094547;
 RA Craik C.S., Choo Q.L., Swift G.H., Quinto C., McDonald R.J.,
 RA Rutter W.J.;
 RT "Structure of two related rat pancreatic trypsin genes.";
 RL J. Biol. Chem. 259:14255-14264(1984).
 RN [2]
 RP SEQUENCE OF 9-246 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Pancreas;
 RX MEDLINE=82265624; PubMed=6896710;
 RA McDonald R.J., Stary S.J., Swift G.H.;
 RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
 sequences of the cloned cDNAs.";
 RL J. Biol. Chem. 257:9724-9732(1982).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.59 ANGSTROMS).
 RX MEDLINE=9135198; PubMed=1881877;
 RA Earnest T., Fauman E., Craik C.S., Stroud R.;
 RT "1.59-A structure of trypsin at 120 K: comparison of low temperature
 and room temperature structures.";
 RL Proteins 10:171-187(1991).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=96214506; PubMed=8634241;
 RA Brinen L.S., Willett W.S., Craik C.S., Fletterick R.J.;
 RT "X-ray structures of a designed binding site in trypsin show metal-
 dependent geometry.";
 RL Biochemistry 35:5999-6009(1996).
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: V01274; CAA24581.1; -;
 CC EMBL: L00131; AAA98517.1; -;
 CC EMBL: L00130; AAA98517.1; JOINED.
 CC PIR: A22657; TRT2.
 DR PDB: 1ANB; 01-APR-97.
 DR PDB: 1ANC; 01-APR-97.
 DR PDB: 1AND; 01-APR-97.
 DR PDB: 1ANE; 01-APR-97.
 DR PDB: 1AMH; 24-DEC-97.
 DR PDB: 1SLU; 11-JUL-96.
 DR PDB: 1SLV; 11-JUL-96.
 DR PDB: 1SLW; 11-JUL-96.
 DR PDB: 1SLX; 11-JUL-96.
 DR PDB: 1DPQ; 07-JUL-97.
 DR PDB: 3TGI; 23-DEC-98.
 DR PDB: 3TGJ; 23-DEC-98.
 DR PDB: 1BRA; 30-APR-94.

DR PDB: 1BRB; 15-JAN-95.
 DR PDB: 1BRC; 31-MAY-94.
 DR PDB: 1CO7; 07-JAN-03.
 DR PDB: 1E2S; 23-JUN-00.
 DR PDB: 1E2U; 23-JUN-00.
 DR PDB: 1F5R; 04-JUL-01.
 DR PDB: 1F7Z; 04-JUL-01.
 DR PDB: 1FY8; 04-JUL-01.
 DR PDB: 1J14; 11-FEB-03.
 DR PDB: 1J15; 11-FEB-03.
 DR PDB: 1J16; 11-FEB-03.
 DR PDB: 1J17; 11-FEB-03.
 DR PDB: 1QL9; 25-AUG-00.
 DR PDB: 1TRM; 15-JUL-93.
 DR PDB: 2TRM; 16-JUL-88.
 DR PDB: 3TCK; 04-JUL-01.
 DR MEROPS; S01.258; -;
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
 KW Multigene family; 3D-structure.
 FT SIGNAL 1 15
 FT PROPEP 16 23 ACTIVATION PEPTIDE.
 FT CHAIN 24 246 TRYPSIN II, ANIONIC.
 FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
 FT ACT_SITE 107 107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 200 200 CHARGE RELAY SYSTEM.
 FT DISULFID 30 160
 FT DISULFID 48 64
 FT DISULFID 132 233
 FT DISULFID 139 206
 FT DISULFID 171 185
 FT DISULFID 196 220
 FT SITE 194 194
 FT CONFLICT 84 84
 FT CONFLICT 88 88
 FT STRAND 25 25
 FT STRAND 28 29
 FT STRAND 32 33
 FT TURN 36 37
 FT STRAND 38 42
 FT STRAND 46 54
 FT TURN 55 56
 FT STRAND 57 60
 FT STRAND 62 64
 FT STRAND 70 73
 FT STRAND 77 77
 FT TURN 78 79
 FT STRAND 86 95
 FT TURN 97 98
 FT TURN 101 103
 FT TURN 105 106
 FT STRAND 109 113
 FT STRAND 120 120
 FT TURN 121 122
 FT STRAND 123 123
 FT STRAND 127 127
 FT TURN 135 136
 FT STRAND 138 143
 FT STRAND 157 157
 FT STRAND 159 164
 FT HELIX 168 174
 FT TURN 176 178
 FT TURN 181 182
 FT STRAND 183 185
 FT STRAND 194 194
 FT TURN 197 198
 FT TURN 200 201
 FT STRAND 203 206

REQUIRED FOR SPECIFICITY (BY SIMILARITY).
 N -> D (IN REF. 1).
 V -> I (IN REF. 1).

```

FT TURN 207 208
FT STRAND 209 216
FT TURN 223 224
FT STRAND 228 231
FT HELIX 232 234
FT HELIX 236 246
SQ SEQUENCE 246 AA; 26228 MW; A8D3630809AE606 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 246;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 19 DDDDK 23

RESULT 13
TRY1_HUMAN
ID TRX1_HUMAN STANDARD; PRT; 247 AA.
AC P07477; Q92955; Q9HAN4; Q9HAN5; Q9HAN6; Q9HAN7;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin I precursor (EC 3.4.21.4) (Cationic trypsinogen).
GN PRSS1 OR TRY1 OR TRP1 OR TRYPI.
OS Homo sapiens (Human).
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=86221712; PubMed=3011602;
RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
RA Matsubara K.;
RT "Cloning, characterization and nucleotide sequences of two cDNAs
RT encoding human pancreatic trypsinogens.";
RL Gene 41:305-310(1986).
[2]
SEQUENCE OF 16-43.
RX MEDLINE=90091010; PubMed=2598466;
RA Kinland M., Russick C., Marks W.H., Borgstroem A.;
RT "Immunoreactive anionic and cationic trypsin in human serum.";
RL Clin. Chm. Acta 184:31-46(1989).
[3]
SEQUENCE OF 68-151 FROM N.A., AND VARIANT HPC HIS-122.
RX MEDLINE=96438847; PubMed=8841182;
RA Whitcomb D.C., Gorry M.C., Preston R.A., Furey W., Sossenheimer M.J.,
RA Ulrich C.D., Martin S.P., Gates L.K. Jr., Amann S.T., Toskes P.P.,
RA Liddle R., McGrath K., Uomo G., Post J.C., Ehrlich G.D.;
RT "Hereditary pancreatitis is caused by a mutation in the cationic
RT trypsinogen gene.";
RL Nat. Genet. 14:141-145(1996).
[4]
SEQUENCE OF 15-67 FROM N.A., AND VARIANT HPC GLY-22.
RX MEDLINE=20389982; PubMed=10930381;
RA Teich N., Ockenga J., Hoffmeister A., Manns M., Mossner J., Keim V.;
RT "Chronic pancreatitis associated with an activation peptide mutation
RT that facilitates trypsin activation.";
RL Gastroenterology 119:461-465(2000).
[5]
SEQUENCE OF 68-151 FROM N.A., AND VARIANTS HPC PRO-104; CYS-116 AND
RX PHE-139.
RA Teich N., Bauer N., Mossner J., Keim V.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[6]
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND PHOSPHORYLATION.
RX MEDLINE=96266496; PubMed=8683601;
RA Gaboriaud C., Serre L., Guy-Crotte O., Forest E.,
RA Fontecilla-Camps J.-C.;
RT "Crystal structure of human trypsin 1: unexpected phosphorylation of
RT Tyr151.";
RL J. Mol. Biol. 259:995-1010(1996).

```

```

RN RP VARIANTS HPC ILE-29 AND HIS-122.
RX MEDLINE=97463797; PubMed=9322498;
RA Gorry M.C., Gabaizedeh D., Furey W., Gates L.K. Jr., Preston R.A.,
RA Aston C.E., Zhang Y., Ulrich C., Ehrlich G.D., Whitcomb D.C.;
RT "Mutations in the cationic trypsinogen gene are associated with
RT recurrent acute and chronic pancreatitis.";
RL Gastroenterology 113:1063-1068(1997).
[8]
RN RP VARIANT HPC ILE-29.
RX MEDLINE=98295575; PubMed=9633818;
RA Teich N., Mossner J., Keim V.;
RT "Mutations of the cationic trypsinogen in hereditary pancreatitis.";
RL Hum. Mutat. 12:39-43(1998).
[9]
RN RP VARIANTS HPC VAL-16 AND HIS-122.
RX MEDLINE=99315544; PubMed=10381903;
RA Witt H., Luck W., Becker M.;
RT "A signal peptide cleavage site mutation in the cationic trypsinogen
RT gene is strongly associated with chronic pancreatitis.";
RL Gastroenterology 117:7-10(1999).
[10]
RN RP VARIANT HPC ARG-23.
RX MEDLINE=99219545; PubMed=10204851;
RA Ferec C., Ragueses O., Salomon R., Roche C., Bernard J.P., Guillot M.,
RA Quere I., Faure C., Mercier B., Audrezet M.P., Guillausseau P.J.,
RA Dupont C., Munnich A., Bignon J.D., Le Bodic L.;
RT "Mutations in the cationic trypsinogen gene and evidence for genetic
RT heterogeneity in hereditary pancreatitis.";
RL J. Med. Genet. 36:228-232(1999).
[11]
RN RP VARIANTS HPC THR-29 AND CYS-122.
RX MEDLINE=21648565; PubMed=11788572;
RA Pfulzer R., Myers E., Applebaum-Shapiro S., Finch R., Ellis I.,
RA Neoptolemos J., Kant J.A., Whitcomb D.C.;
RT "Novel cationic trypsinogen (PRSS1) N29I and R122C mutations cause
RT autosomal dominant hereditary pancreatitis.";
RL Gut 50:271-272(2002).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MASS SPECTROMETRY: MW=24348; MW_ERR=2; METHOD=Electrospray;
CC RANGE=24-247.
CC -1- DISEASE: Defects in PRSS1 are a cause of hereditary pancreatitis,
CC (HPC or HP); also known as chronic pancreatitis (CP). HPC is an
CC autosomal dominant disease characterized by the presence of
CC calculi in pancreatic ducts. It causes severe abdominal pain
CC attacks.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M22612; AAA61231.1; -
DR EMBL; U70137; AAC50728.1; -
DR EMBL; AF314534; AAG30943.1; -
DR EMBL; AF315309; AAG30947.1; -
DR EMBL; AF315310; AAG30948.1; -
DR EMBL; AF315311; AAG30949.1; -
DR PIR; A25852; A25852.
DR PDB; 1TRN; 03-JUN-95.
DR PDB; 1FXI; 17-JUN-98.
DR MEROPS; S01.151; -
DR Genew; HGNC:9475; PRSS1.
DR MIM; 276000; -
DR MIM; 167800; -
DR GO; GO:0004295; F:trypsin activity; TAS.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser-protease_Try.

```



```

DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; Phosphorylation; 3D-structure; Disease mutation.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 247
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT DISULFID 30 160
FT DISULFID 48 64
FT DISULFID 139 206
FT DISULFID 171 185
FT DISULFID 196 220
FT MOD_RES 154 154
FT SITE 194 194
FT VARIANT 16 16
FT VARIANT 22 22
FT VARIANT 23 23
FT VARIANT 29 29
FT VARIANT 29 29
FT VARIANT 104 104
FT VARIANT 116 116
FT VARIANT 122 122
FT VARIANT 122 122
FT VARIANT 139 139
FT VARIANT 139 139
FT SEQUENCE 247 AA; 26558 MW; DD49A487B8062813 CRC64;
Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 19 DDDDK 23
RESULT 14
TRY2_BOVIN STANDARD; PRT; 247 AA.
AC Q29463;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin, anionic precursor (EC 3.4.21.4).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoides;

```

```

OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein-Friesian; TISSUE=Pancreas;
RX MEDLINE=91005383; PubMed=1701147;
RT le Huerou I., Wicker C., Guillemeau P., Toullec R., Puigserver A.;
RT "Isolation and nucleotide sequence of cDNA clone for bovine
RT pancreatic anionic trypsinogen. Structural identity within the
RT trypsin family."
RL Eur. J. Biochem. 193;767-773(1990).
RC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54703; CAA38513.1; -.
DR PIR; S13813; S13813.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.258; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 247
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT DISULFID 30 160
FT DISULFID 48 64
FT DISULFID 132 233
FT DISULFID 139 206
FT DISULFID 171 185
FT DISULFID 196 220
FT SITE 194 194
FT SEQUENCE 247 AA; 26289 MW; 50A070495A7731DB CRC64;
Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 19 DDDDK 23
RESULT 15
TRY2_CANFA STANDARD; PRT; 247 AA.
AC P06872;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trypsin, anionic precursor (EC 3.4.21.4).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284628; PubMed=3841794;

```

RA Pinsky S.D., Laforge K.S., Scheele G.;
 RT "Differential regulation of trypsinogen mRNA translation: full-length
 RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
 RT in the dog pancreas.";
 RL Mol. Cell. Biol. 5:2669-2676(1985).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11589; AAA30899.1; -;
 DR PIR; A26273; TRDG.
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.258; -;
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
 KW Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 23
 FT CHAIN 24 247
 FT ACT_SITE 63 63
 FT ACT_SITE 107 107
 FT ACT_SITE 200 200
 FT DISULFID 30 160
 FT DISULFID 48 64
 FT DISULFID 132 233
 FT DISULFID 139 206
 FT DISULFID 171 185
 FT DISULFID 196 220
 FT SITE 194 194
 FT SEQUENCE 247 AA; 26423 MW; 374E9D31D6D88EAF CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 247;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DDDDK 5
 Db 19 DDDDK 23
 RESULT 16
 TRY2_HUMAN
 ID TRY2_HUMAN STANDARD; PRT; 247 AA.
 AC P07478;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trypsin II precursor (EC 3.4.21.4) (Anionic trypsinogen).
 GN PRS2 OR TRY2 OR TRY2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86221712; PubMed=3011602;
 RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
 RA Matsubara K.;
 RT "Cloning, characterization and nucleotide sequences of two cDNAs
 RT encoding human pancreatic trypsinogens.";

RL Gene 41:305-310(1986).
 RN [2]
 RP SEQUENCE OF 16-49.
 RX MEDLINE=90091010; PubMed=2598466;
 RA Kinsland M., Russick C., Marks W.H., Borgstrom A.;
 RT "Immunoreactive anionic and cationic trypsin in human serum.";
 RL Clin. Chim. Acta 184:31-46(1989).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M27602; AAA61232.1; -;
 DR PIR; B25852; B25852.
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.258; -;
 DR Genew; HGNC:9483; PRSS2.
 DR MIM; 601564; -;
 DR GO; GO:0004295; F:trypsin activity; TAS.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
 KW Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 23
 FT CHAIN 24 247
 FT ACT_SITE 63 63
 FT ACT_SITE 107 107
 FT ACT_SITE 200 200
 FT DISULFID 30 160
 FT DISULFID 48 64
 FT DISULFID 171 185
 FT DISULFID 196 220
 FT SITE 194 194
 FT SEQUENCE 247 AA; 26488 MW; 82B0F41EB8E3D5DB CRC64;
 Query Match 100.0%; Score 29; DB 1; Length 247;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DDDDK 5
 Db 19 DDDDK 23
 RESULT 17
 TRY3_RAT
 ID TRY3_RAT STANDARD; PRT; 247 AA.
 AC P08426;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trypsin III, cationic precursor (EC 3.4.21.4) (Pretrypsinogen III).
 GN TRY3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87271609; PubMed=3607011;
 RA Fletcher T.S., Alhadeff M., Craik C.S., Largman C.;

```

RT "Isolation and characterization of a cDNA encoding rat cationic
RT trypsinogen.";
RL Biochemistry 26:3081-3086(1987).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16624; AAA41985.1; -
DR PIR: A27547; A27547.
DR HSSP: P00763; LDPO.
DR MEROPS: S01.056; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP-SPC; 1.
DR PROSITE: PS02440; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase: Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 24 ACTIVATION PEPTIDE.
FT CHAIN 25 247 TRYPSIN III, CATIONIC.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 161 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 234 BY SIMILARITY.
FT DISULFID 140 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 197 221 BY SIMILARITY.
FT SITE 195 195 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26269 MW; D74892BAA364E4A8 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 20 DDDDK 24

RESULT 18
SN29_ARATH
ID SN29_ARATH STANDARD; PRT; 251 AA.
AC Q9SD96;
RT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE SNAP25 homologous protein SNAP29 (AtSNAP29) (Synaptosomal-associated
DE protein SNAP25-like 2).
GN SNAP29 OR AT5G07880 OR XM12.11 OR F13C24.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukami M.,

```

```

RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones.";
RL DNA Res. 4:215-230(1997).
CC -1- FUNCTION: Vesicle trafficking protein that functions in the
CC secretory pathway (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SNAP-25 FAMILY.
CC -1- SIMILARITY: Contains 1 t-SNARE coiled-coil homology domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL133421; CAB62600.1; -
DR EMBL: AB005249; BAB09952.1; -
DR PIR: T45613; T45613.
DR InterPro: IPR000727; T-SNARE.
DR SMART: SM00397; t-SNARE; 2.
DR PROSITE: PS50192; T-SNARE; 1.
DR Transport; Protein transport; Membrane; Coiled coil; Multigene family.
KW DOMAIN 186 248 T-SNARE COILED-COIL HOMOLOGY.
FT DOMAIN 186 248
SQ SEQUENCE 251 AA; 28308 MW; 406BE52160B205DF CRC64;

Query Match 100.0%; Score 29; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 29 DDDDK 33

RESULT 19
ASP_PLAFS
ID ASP_PLAFS STANDARD; PRT; 253 AA.
AC P13825;
RT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartic acid-rich protein precursor.
OS Plasmodium falciparum (isolate fcm17 / Senegal).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5845;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87270765; PubMed=3038111;
RA Lenstra R., D'Auriol L., Andrieu B., le Bras J., Galibert F.;
RT "Cloning and sequencing of Plasmodium falciparum DNA fragments
RT containing repetitive regions potentially coding for histidine-rich
RT proteins. Identification of two overlapping reading frames.";
RL Biochem. Biophys. Res. Commun. 146:368-377(1987).
CC -1- MISCELLANEOUS: THIS PROTEIN IS CODED ON THE REVERSE STRAND OF AN
CC HISTIDINE-RICH PROTEIN.
CC -1- SIMILARITY: SOME SIMILARITY WITH HUMAN SET PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17028; AAA29620.1; -
DR PIR: B29653; B29653.
DR InterPro: IPR002164; NAP_family.
DR Pfam: PF00956; NAP; 1.
KW Malaria; Antigen; Signal.

```

FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 253 ASPARTIC ACID-RICH PROTEIN.
 FT DOMAIN 204 253 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 253 AA; 30248 MW; 83E585DE074B8504 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 253;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 232 DDDDK 236

RESULT 20
 LI39_CAEEL STANDARD; PRT; 253 AA.
 AC P34684;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox protein lin-39.
 GN LIN-39 OR CEH-15 OR C07H6.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA MEDLINE=93327430; PubMed=8101475;
 RX Clark S.G., Chisholm A.D., Horvitz H.R.;
 RT "Control of cell fates in the central body region of C. elegans by
 the homeobox gene lin-39";
 RL Cell 74:43-55(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93327429; PubMed=8101474;
 RA Wang B.B., Mueller-Immergluck M.M., Austin J., Robinson N.T.,
 RA Chisholm A.D., Kenyon C.;
 RT "A homeotic gene cluster patterns the anteroposterior body axis of C.
 elegans";
 RL Cell 74:29-42(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Macri C., Vaudin M.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CONTROLS THE MIGRATION OF NEUROBLASTS AND THE
 SUBSEQUENT MID-BODY REGION-SPECIFIC DEVELOPMENT.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
 CC "DEFORMED" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; L19639; AAC37168.1; -;
 CC EMBL; L19248; AAB04137.1; -;
 CC EMBL; AC006605; AAK85445.1; -;
 CC PIR; B40722; B40722.
 CC HSSP; P02833; ISAN.
 CC WormPep; C07H6.7; CE03975.
 CC TRANSFAC; T03369; -;
 CC InterPro; IPR001827; Antennapedia.
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00025; ANTENNAPEDIA.

DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.
 KW Homeobox; DNA-binding; Nuclear protein; Developmental protein.
 FT DOMAIN 20 36 POLY-SER.
 FT DOMAIN 121 129 ANTP-ASP.
 FT SITE 148 153 ANTP-TYPE HEXAPEPTIDE.
 FT DNA_BIND 165 224 HOMEBOX.
 FT DOMAIN 228 248 PRO-RICH.
 SQ SEQUENCE 253 AA; 27814 MW; 6B59DE9568B834A1 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 253;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 |||||
 Db 122 DDDDK 126

RESULT 21
 SN30_ARATH STANDARD; PRT; 263 AA.
 ID SN30_ARATH
 AC Q9LMG8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative SNAP25 homologous protein SNAP30 (AtSNAP30) (Synaptosomal-
 associated protein SNAP25-like 3).
 GN SNAP30 OR ATG13890 OR F16A14.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eumids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RL Nature 408:816-820(2000).
 CC -!- FUNCTION: Vesicle trafficking protein that functions in the
 CC secretory pathway (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SNAP-25 FAMILY.
 CC -!- SIMILARITY: Contains 1 t-SNARE coiled-coil homology domain.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC EMBL; AC068197; AAF79396.1; -.
DR PIR; A86272; A86272.
DR InterPro; IPR000727; T_SNARE.
DR SMART; SM00397; T_SNARE; 2.
DR PROSITE; PS0192; T_SNARE; 1.
KW Hypothetical protein; Transport; Protein transport; Membrane;
KW Coiled coil; Multigene family.
FT DOMAIN 198 260 T-SNARE COILED-COIL HOMOLOG.
SQ SEQUENCE 263 AA; 29075 MW; D55EE680CFE02C4 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 43 DDDDK 47

RESULT 22
YG5W_YEAST
ID YG5W_YEAST STANDARD; PRT; 271 AA.
AC P53335;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 31.3 kDa protein in TAF145-YOR1 intergenic region.
GN YGR280C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=288c / FY1679;
RX MEDLINE=97245295; PubMed=90900054;
RA Volckaert G., Voet M., Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MAL1 locus reveals 15 complete open reading frames, including
RT ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";
RL Yeast 13:251-259(1997).
CC -!- SIMILARITY: Contains 1 G-patch domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73065; CAA97311.1; -.
DR PIR; S64615; S64615.
DR SGD; S0003512; YGR280C.
DR GO; GO:0006365; P:35S primary transcript processing; IMP.
DR GO; GO:0016180; P:snRNA processing; IMP.
DR InterPro; IPR000467; G_patch.
DR Pfam; PF01585; G_patch; 1.
DR SMART; SM00443; G_patch; 1.
DR PROSITE; PS0174; G_PATCH; 1.
KW Hypothetical protein.
FT DOMAIN 25 72 G-PATCH.
SQ SEQUENCE 271 AA; 31311 MW; D5AB6EB2C129922D CRC64;

Query Match 100.0%; Score 29; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 163 DDDDK 167

```

```

RESULT 23
SPBP_RAT
ID SPBP_RAT STANDARD; PRT; 279 AA.
AC P08723;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prostatic spermine-binding protein precursor (SBP).
GN SBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87137538; PubMed=3818623;
RA Chang C., Saltzman A.G., Hiipakka R.A., Huang I.-Y., Liao S.;
RT "Prostatic spermine-binding protein. Cloning and nucleotide sequence
RT of cDNA, amino acid sequence, and androgenic control of mRNA level.";
RL J. Biol. Chem. 262:2826-2831(1987).
RN [2]
RP SEQUENCE, AND REVISIONS.
RX MEDLINE=89000602; PubMed=3166977;
RA Anderegg R.J., Carr S.A., Huang I.-Y., Hiipakka R.A., Chang C.,
RA Liao S.;
RT "Correction of the cDNA-derived protein sequence of prostatic
RT spermine binding protein: pivotal role of tandem mass spectrometry in
RT sequence analysis.";
RL Biochemistry 27:4214-4221(1988).
CC -!- FUNCTION: SPERMINE-BINDING PROTEIN IS AN ANDROGEN REGULATED
CC VENTRAL PROSTATE GLYCOPROTEIN THAT BINDS VARIOUS POLYAMINES.
CC -!- TISSUE SPECIFICITY: Prostate.
CC -!- SIMILARITY: TO MOUSE SBP.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02675; AAA42113.1; ALT-SEQ.
DR InterPro; IPR001229; Jacalin_lectin.
DR Pfam; PF01419; Jacalin; 1.
KW Glycoprotein; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 17
FT CHAIN 18 279 PROSTATIC SPERMINE-BINDING PROTEIN.
FT MOD_RES 18 18 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 62 62 N-LINKED (GLCNAC).
FT DOMAIN 154 279 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 279 AA; 31080 MW; 3BEB01A02A517A65 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 194 DDDDK 198

RESULT 24
YGLB_YEAST
ID YGLB_YEAST STANDARD; PRT; 297 AA.
AC P53210;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 34.7 kDa protein in MSB2-UGA1 intergenic region.
GN YGR017W.

```

```
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomycetes cerevisiae
RL chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -!- SIMILARITY: TO S.POMBE SPBC725.03.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72802; CAA97000.1; -
DR PIR; S64308; S64308
DR SGD; S0003249; YGR017W.
KW Hypothetical protein.
SQ SEQUENCE 297 AA; 34679 MW; 2CA638D9F734C6C1 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 156 DDDDK 160

RESULT 25
TRY3_HUMAN STANDARD; PRT; 304 AA.
AC P35030; P15951; Q15665; Q9UQV3;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Trypsin III precursor (EC 3.4.21.4) (Brain trypsinogen)
DE (Mesotrypsinogen) (Trypsin IV).
GN PRSS3 OR TRY3 OR PRSS4 OR TRY4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
RC TISSUE=Brain;
RX MEDLINE=94123994; PubMed=8294000;
RA Wiegand U., Corbach S., Minn A., Kang J., Mueller-Hill B.;
RT "Cloning of the cDNA encoding human brain trypsinogen and
RL characterization of its product."
RL Gene 136:167-175(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM C).
RC TISSUE=Pancreas;
RX MEDLINE=90221895; PubMed=2326201;
RA Tani T., Kawashima I., Mita K., Takiguchi Y.;
RT "Nucleotide sequence of the human pancreatic trypsinogen III cDNA.";
RL Nucleic Acids Res. 18:1631-1631(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Fukuoka S.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) (ISOFORM A).
MEDLINE=21686310; PubMed=11827488;
```

```
RA Katona G., Berglund G.I., Hajdu J., Graf L., Szilagyi L.;
RT "Crystal structure reveals basis for the inhibitor resistance of human
RL brain trypsin.";
RL J. Mol. Biol. 315:1209-1218(2002).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A;
CC IsoId=P35030-1; Sequence=Displayed;
CC Name=B;
CC IsoId=P35030-2; Sequence=VSP_005409;
CC Name=C;
CC IsoId=P35030-3; Sequence=VSP_005410;
CC -!- TISSUE SPECIFICITY: Pancreas and brain.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72781; CAB58178.1; -
DR EMBL; X71345; CAA50484.1; -
DR EMBL; X15505; CAA33527.1; -
DR EMBL; D45417; BAA08257.1; -
DR PIR; S33496; S33496
DR PDB; 1H4W; 15-FEB-02.
DR MEROPS; S01.174; -
DR Genew; HGNC:9486; PRSS3.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0004295; F:trypsin activity; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; FALSE_NEG.
KW Hydrolyase; Serine protease; Digestion; Zymogen; Signal;
KW Multigene family; Alternative splicing; 3D-structure.
FT SIGNAL 1 ? POTENTIAL.
FT PROPEP 7 80 ACTIVATION PEPTIDE.
FT CHAIN 81 304 TRYPsin III.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM.
FT ACT_SITE 257 257 CHARGE RELAY SYSTEM.
FT DISULFID 87 217
FT DISULFID 105 121
FT DISULFID 196 263
FT DISULFID 228 242
FT DISULFID 253 277
FT SITE 251 251
FT VARSPPLIC 1 45
FT -----
FT VARSPPLIC 1 70
FT -----
FT CONFLICT 89 89 MISSING (IN REF. 1; CAA50484).
FT CONFLICT 224 225 TQ -> RE (IN REF. 2).
FT CONFLICT 232 232 Y -> C (IN REF. 2).
FT CONFLICT 253 254 CQ -> WK (IN REF. 2).
SQ SEQUENCE 304 AA; 32499 MW; 4C4316C31F1D0FFC CRC64;

Query Match 100.0%; Score 29; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 94;
```

```

Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 DDDDK 5
DB      76 DDDDK 80

RESULT 26
VAS1_VACCV
ID VAS1_VACCV      STANDARD;      PRT;      334 AA.
AC P21069;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein A51.
GN A51R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus.'";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35027; AAA48185.1; -
CC PIR; H42522; H42522.
CC InterPro: IPR007032; Pox_A51.
CC Pfam: PF04948; Pox_A51; 1.
SQ SEQUENCE 334 AA; 37723 MW; 508174C65ADEDB54 CRC64;

Query Match      100.0%; Score 29; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 DDDDK 5
DB      89 DDDDK 93

RESULT 27
VAS1_VACCV
ID VAS1_VACCV      STANDARD;      PRT;      334 AA.
AC Q01219;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Protein A51.
GN A51R OR SALF14R.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91259063; PubMed=2045793;

```

```

RA Smith G.L., Chan Y.S., Howard S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT the right inverted terminal repeat.";
RL J. Gen. Virol. 72:1349-1376(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11079; BAA01825.1; -
CC PIR; JQ1789; JQ1789.
CC InterPro: IPR007032; Pox_A51.
CC Pfam: PF04948; Pox_A51; 1.
SQ SEQUENCE 334 AA; 37731 MW; AF82B07B8BC21544 CRC64;

Query Match      100.0%; Score 29; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 DDDDK 5
DB      89 DDDDK 93

RESULT 28
VAS1_VARV
ID VAS1_VARV      STANDARD;      PRT;      334 AA.
AC P33858;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Protein A51.
GN A51R OR J5R.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=India-1967 / Isolate Ind3;
RX MEDLINE=92209372; PubMed=1666548;
RA Shchelkunov S.N., Marennikova S.S., Totmenin A.V., Blinov V.M.,
RA Chizhikov V.E., Gutorov V.V., Saifonov P.F., Pozdnyakov S.G.,
RA Shelukhina E.M., Gashnikov P.V., Anjaparidze O.G., Sandakhchiev L.S.;
RT "Creation of a clone library of fragments from the natural variola
RT virus and study of the structural and functional organization of
RT viral genes from a circle of hosts.";
RL Dokl. Akad. Nauk SSSR 321:402-406(1991).
RN [2]
RP COMPLETE GENOME.
RC STRAIN=India-1967 / Isolate Ind3;
RX MEDLINE=93202281; PubMed=8384129;
RA Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Harvey;
RX MEDLINE=93057361; PubMed=1331292;
RA Aguado B., Selmes I.P., Smith G.L.;
RT "Nucleotide sequence of 21.8 kbp of variola major virus strain Harvey
RT and comparison with vaccinia virus.";
RL J. Gen. Virol. 73:2887-2902(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bangladesh-1975;
RX MEDLINE=94088747; PubMed=8264798;
RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,

```

RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Many B.W.J.,
RA Venter C.J.;
RT "Potential virulence determinants in terminal regions of variola
RL smallpox virus genome.";
RL Nature 366:748-751(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69198; CAA49104.1; -;
DR EMBL; X67118; CAA47546.1; -;
DR EMBL; L22579; AAA60906.1; -;
DR PIR; B72171; B72171.
DR PIR; S46845; S46845.
DR PIR; T28596; T28596.
DR InterPro; IPR007032; Pox_A51.
DR Pfam; PF04948; Pox_A51; 1.
SQ SEQUENCE 334 AA; 37641 MW; A01D5461A0119C81 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 89 DDDDK 93

RESULT 29
ID CWG2_SCHPO STANDARD; PRT; 355 AA.
AC P32434;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Type I protein geranyltransferase beta subunit (EC 2.5.1.-)
DE beta) (PGST)
DE CWG2 OR SPAC2E1P5.04C.
GN Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94085400; PubMed=8262067;
RA Diaz M., Sanchez Y., Bennett T., Dun C.R., Godoy C., Tamanoi F.,
RA Duran A., Perez P.;
RA "The Schizosaccharomyces pombe cwg2+ gene codes for the beta subunit
RT of a geranyltransferase type I required for beta-glucan
RT synthesis.";
RL EMBO J. 12:5245-5254(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scourtos J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovskii G.V., Ussery D., Barrell B.G., Nurse P.;
RT The genome sequence of Schizosaccharomyces pombe.;
RL Nature 415:871-880(2002).
CC 1- FUNCTION: CATALYZES THE TRANSFER OF A GERANYL-GERANYL MOIETY FROM
CC GERANYLGERANYL PYROPHOSPHATE TO PROTEINS HAVING THE C-TERMINAL C-
CC A-A-L. IN PARTICULAR IT MODIFIES THE GTP-BINDING COMPONENT OF THE
CC (1-3)BETA-D-GLUCAN SYNTHASE.
CC 1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC 1- SUBUNIT: Heterodimer of an alpha and a beta subunit.
CC 1- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT
CC FAMILY
CC 1- SIMILARITY: Contains 4 PFTB repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z12155; CAA78143.1; -;
DR EMBL; AL163071; CAB86347.1; -;
DR PIR; S41686; S41686.
DR HSP; Q02293; IFT1.
DR GenedB_SPombe; SPAC2E1P5.04c; -;
DR InterPro; IPR001330; Prenyltrans.
DR Pfam; PF00432; prenyltrans; 4.
KW Transferase; Prenyltransferase; Repeat; Zinc.
FT REPEAT 119 163 PFTB 1.
FT REPEAT 169 210 PFTB 2.
FT REPEAT 234 275 PFTB 3.
FT REPEAT 282 324 PFTB 4.
FT METAL 260 260 ZINC (BY SIMILARITY).
FT METAL 262 262 ZINC (BY SIMILARITY).
FT METAL 312 312 ZINC (BY SIMILARITY).
SQ SEQUENCE 355 AA; 40023 MW; 811A3CBFF0067071 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 51 DDDDK 55

RESULT 30
ID YBD2_YEAST STANDARD; PRT; 381 AA.
AC P38199;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 41.7 kDa protein in RIB1-AAC2 intergenic region.
GN YBL032W OR YBL0418.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;


```

RN SEQUENCE FROM N.A.
RP STRAIN-S288C;
RX MEDLINE=95242843; PubMed=7725803;
RA van Dyck L., Joniaux J.-L., Barreiros T.D.M., Kline K., Goffeau A.;
RT "Analysis of a 17.4 kb DNA segment of yeast chromosome II
RT encompassing the ribosomal protein L19 as well as proteins with
RT homologies to components of the hnRNP and snRNP complexes and to the
RT human proliferation-associated p120 antigen.";
RL Yeast 10:1663-1673(1994).
CC -1- FUNCTION: COULD BE A MEMBER OF THE HNRNP COMPLEXES.
CC -1- SIMILARITY: Contains 4 KH domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77291; CAA54496.1; -.
DR EMBL; Z35793; CAA84852.1; -.
DR PIR; S45766; S45766.
DR SGD; S0000128; YBL032W.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005696; C:telomere; IDA.
DR GO; GO:0003729; F:mRNA binding activity; IDA.
DR GO; GO:0008298; P:mRNA localization, intracellular; IMP.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; IMP.
DR InterPro; IPR004087; KH_dom.
DR IntronPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS00084; KH_TYPE_1; 3.
KW Hypothetical protein; RNA-binding; Repeat.
FT DOMAIN 43 81
FT DOMAIN 156 191
FT DOMAIN 192 230
FT DOMAIN 258 296
FT DOMAIN 381 AA; 41683 MW; OA8752961EB77DC9 CRC64;
SQ SEQUENCE 381 AA; 41683 MW; OA8752961EB77DC9 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 298 DDDDK 302

RESULT 31
METK_BIFLO
ID METK_BIFLO STANDARD; PRT; 406 AA.
AC Q8G3H4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE adenosyltransferase) (AdoMet synthetase) (MAT).
GN METK OR BL1786.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCC 2705;
RA MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwaan M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation

```

```

RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -1- FUNCTION: Catalyzes the formation of S-adenosylmethionine from
CC methionine and ATP. The overall synthetic reaction is composed of
CC two sequential steps, AdoMet formation and the subsequent
CC triphosphate hydrolysis which occurs prior to release of
CC AdoMet from the enzyme (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -1- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and
CC 1 potassium ion per subunit (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ADOLET SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014812; AAN25569.1; -.
DR HAMAP; MF_00086; -. 1.
DR Pfam; PF00438; S-AdoMet_synt; 1.
DR Pfam; PF02772; S-AdoMet_synt2; 1.
DR Pfam; PF02773; S-AdoMet_synt3; 1.
DR TIGRfams; TIGR01034; metk; 1.
DR PROSITE; PS00376; ADOLET SYNTHETASE_1; 1.
DR PROSITE; PS00377; ADOLET SYNTHETASE_2; 1.
KW Transferase; One-carbon metabolism; ATP-binding; Magnesium; Potassium;
KW Metal-binding; Complete proteome.
FT NP_BIND 279 286 ATP (POTENTIAL).
FT METAL 19 19 MAGNESIUM (BY SIMILARITY).
FT METAL 45 45 POTASSIUM (BY SIMILARITY).
FT METAL 283 283 POTASSIUM (BY SIMILARITY).
FT METAL 291 291 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 406 AA; 43775 MW; ABFF474AE188C26A CRC64;

Query Match 100.0%; Score 29; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 188 DDDDK 192

RESULT 32
QSEC_HAEIN
ID QSEC_HAEIN STANDARD; PRT; 451 AA.
AC P45336;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensor protein qsec (EC 2.7.3.-).
GN QSEC OR H11707.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

```

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae";
RL Electrophoresis 21:411-429(2000)
CC -!- FUNCTION: Member of a two-component regulatory system qseB/qseC.
CC May activate qseB by phosphorylation (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -!- SIMILARITY: Contains 1 HAM domain.
CC -!- SIMILARITY: Contains 1 histidine kinase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32843; AAC23353.1; -;
DR PIR; E64137; E64137.
DR TIGR; H11707; -;
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR004358; Bact_sens_pr_c.
DR InterPro; IPR003660; HAMK.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR005467; His_kinase.
DR Pfam; PF00672; HAMK; 1.
DR Pfam; PF02518; HAMKase_c; 1.
DR Pfam; PF05512; Hiska; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HAMKase_c; 1.
DR SMART; SM00388; Hiska; 1.
DR PROSITE; PS00885; HAMK; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
KW Sensory transduction; transferase; Kinase; Transmembrane;
KW Inner membrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 9
FT TRANSMEM 10 31
FT DOMAIN 32 162
FT PERIPLASMIC (POTENTIAL).
FT TRANSMEM 163 180
FT POTENTIAL.
FT DOMAIN 181 451
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 241 233
FT HAMK.
FT DOMAIN 241 451
FT HISTIDINE KINASE.
FT MOD_RES 244 244
FT PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 451 AA; 51271 MW; 00C86B648128D383 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 122 DDDDK 126

RESULT 33
MK11 YEAST
ID MK11 YEAST STANDARD; PRT; 468 AA.
AC P20484;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE MAK11 protein.

GN MAK11 OR YKL021C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88087288; PubMed=2826479;
RA Icho T., Wickner R.B.;
RT "The MAK11 protein is essential for cell growth and replication of M
RT double-stranded RNA and is apparently a membrane-associated
RT protein.";
RL J. Biol. Chem. 263:1467-1475(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Rieger M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR CELL GROWTH AND REPLICATION OF M1 DOUBLE-
CC STRANDED RNA.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- SIMILARITY: Contains 6 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03506; AAA34750.1; -;
DR EMBL; Z28021; CAA81856.1; -;
DR PIR; A29938; A29938.
DR SGD; S0001504; MAK11.
DR GO; GO:0005624; C:membrane fraction; IDA.
DR GO; GO:0042273; P:ribosomal large subunit biogenesis; IMP.
DR GO; GO:0019048; P:virus-host interaction; IMP.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat; Membrane.
FT REPEAT 104 132
FT REPEAT 144 189
FT REPEAT 201 231
FT REPEAT 243 275
FT REPEAT 292 321
FT REPEAT 352 384
FT REPEAT 440 462
FT DOMAIN 440 462
FT LYS-RICH (BASIC).
SQ SEQUENCE 468 AA; 53672 MW; 7489AE9139D99FAC CRC64;

Query Match 100.0%; Score 29; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 337 DDDDK 341

RESULT 34
VE2 HPV20
ID VE2 HPV20 STANDARD; PRT; 497 AA.
AC P50766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Regulatory protein E2.
GN E2.

OS Human papillomavirus type 20.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus
 OX NCBI_TaxID=31547;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Delius H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U31778; AAA79390.1; -;
 DR HSSP; P03122; 2BOP.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 SQ SEQUENCE 497 AA; 56120 MW; 497CA228E3537CE5 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 497;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 141 DDDDK 145

RESULT 35
 VE2 HPV25
 ID VE2 HPV25 STANDARD; PRT; 502 AA.
 AC P36787;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.

OS Human papillomavirus type 25.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10609;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Delius H., Hofmann B.;
 RL "Primer-directed sequencing of human papillomavirus types."; Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACGNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; X74471; CAA52527.1; -;
 DR PIR; S36494; S36494.
 DR HSSP; P03122; 2BOP.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 502 AA; 56724 MW; FC0115988D7C8DB6 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 502;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 |||||
 Db 141 DDDDK 145

RESULT 36
 CTK1 YEAST
 ID CTK1 YEAST STANDARD; PRT; 528 AA.
 AC O03957;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE CTD kinase alpha subunit (EC 2.7.1.-) (CTD kinase 58 kDa subunit)
 DE (CTDK-I alpha subunit).
 GN CTK1 OR YKL139W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92314702; PubMed=1820212;
 RA Lee J.M., Greenleaf A.L.;
 RT "CTD kinase large subunit is encoded by CTK1, a gene required for
 RT normal growth of Saccharomyces cerevisiae."; Gene Expr. 1:149-167(1991).
 RN [2]
 RP SEQUENCE FROM N.A.

RA Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CTDK-I HYPERPHOSPHORYLATES THE CARBOXYL-TERMINAL REPEAT
 CC DOMAIN (CTD) OF RNA POLYMERASE II LARGEST SUBUNIT. THIS PROTEIN IS
 CC THE CATALYTIC SUBUNIT.
 CC -!- SUBUNIT: CONSISTS OF THREE SUBUNITS (ALPHA, BETA, GAMMA) OF 58,
 CC 38, AND 32 kDa, RESPECTIVELY.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CDC2/CDKX SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M69024; AAC41642.1; -;
 DR EMBL; Z28139; CAA81980.1; -;
 DR PIR; S32593; S32593.
 DR HSSP; P24941; LHCL.
 DR SGD; S0001622; CTK1.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0004693; F:cyclin-dependent protein kinase activity; IDA.
 DR InterPro; IPR000719; Prot_kinase.


```

CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000714; AAC07029.1; -.
CC PIR: B70381; B70381.
CC MEROPS: M41.009; -.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003959; AAA_ATPase_cent.
CC InterPro: IPR003960; AAA_sub.
CC InterPro: IPR005936; FtsH.
CC InterPro: IPR000642; Peptidase_M41.
CC Pfam: PF00004; AAA; 1.
CC SMART: SM00382; AAA; 1.
CC TIGRfams: TIGR01241; FtsH_fam; 1.
CC PROSITE: PS00674; AAA; 1.
CC Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;
CC Zinc; Complete proteome.
CC DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 6 26 POTENTIAL.
CC DOMAIN 27 101 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 102 122 POTENTIAL.
CC DOMAIN 123 634 CYTOPLASMIC (POTENTIAL).
CC NP_BIND 195 202 ATP (POTENTIAL).
CC METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 419 419 BY SIMILARITY.
CC METAL 422 422 ZINC (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 634 AA; 70686 MW; 320D4A6B3F58AA53 CRC64;
Query Match 100.0%; Score 29; DB 1; Length 634;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
DB 430 DDDDK 434
RESULT 40
PDR4_YEAST
ID PDR4_YEAST STANDARD; PRT; 650 AA.
AC P19880; P22631;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcriptional activator PDR4 (YAP-1 protein).
GN PDR4 OR YAP1 OR SNQ3 OR PAR1 OR YML007W OR YN571.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89252807; PubMed=2542125;
RA Moye-Rowley W.S., Harshan K.D., Parker C.S.;
RT "Yeast YAP1 encodes a novel form of the jun family of transcriptional
RT activator proteins.";
RL Genes Dev. 3:283-292(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91285426; PubMed=2060792;
RA Hussain M., Lenard J.;
RT "Characterization of PDR4, a Saccharomyces cerevisiae gene that
RT confers pleiotropic drug resistance in high-copy number: identity
RT with YAP1, encoding a transcriptional activator.";

```

```

RL Gene 101:149-152(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347405; PubMed=1878996;
RA Hertle K., Haase E., Brendel M.;
RT "The SNQ3 gene of Saccharomyces cerevisiae confers hyper-resistance
RT to several functionally unrelated chemicals.";
RN Curr. Genet. 19:429-433(1991).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91364700; PubMed=1889413;
RA Schnell N., Entian K.-D.;
RT "Identification and characterization of a Saccharomyces cerevisiae
RT gene (PAR1) conferring resistance to iron chelators.";
RN Eur. J. Biochem. 200:487-493(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Tagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RN Nature 387:90-93(1997).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=92405220; PubMed=1525853;
RA Schnell N., Krebs B., Entian K.-D.;
RT "The PAR1 (YAP1/SNQ3) gene of Saccharomyces cerevisiae, a c-jun
RT homologue, is involved in oxygen metabolism.";
RN Curr. Genet. 21:269-273(1992).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR INVOLVED IN THE GENE
CC REGULATION OF CERTAIN OXYGEN DETOXIFICATION ENZYMES. RECOGNIZES
CC THE 5'-TGACTCA-3' SEQUENCE ELEMENT COMMON IN GCN4/AP-1 SITES. ITS
CC OVERPRODUCTION CAUSES CELLS TO ACQUIRE A PLEIOTROPIC DRUG-
CC RESISTANCE AND TO BE ABLE TO TOLERATE NORMALLY TOXIC LEVELS OF
CC CADMIUM, IRON CHELATORS AND ZINC. REGULATES THE EXPRESSION OF THE
CC THIOREDOXIN (YRX2) AND THE YCF1 GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the bzip family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X58693; CAA41536.1; -.
CC EMBL: X53830; CAA37827.1; -.
CC EMBL: X63268; CAA44917.1; -.
CC EMBL: Z49810; CAA89945.1; -.
CC PIR: S16706; S16706.
CC TRANSFAC: T00028; -.
CC SGD: S0004466; YAP1.
CC GO: GO:0005737; C:cytoplasm; IDA.
CC GO: GO:0005634; C:nucleus; IDA.
CC InterPro: IPR004827; TF_bZIP.
CC SMART: SM00338; BRLZ; 1.
CC PROSITE: PS50217; BZIP; 1.
CC PROSITE: PS00036; BZIP_BASIC; 1.
CC Transcription regulation; Activator; DNA-binding; Nuclear protein;
CC Cadmium resistance.
CC DNA_BIND 67 90 BASIC MOTIF.
CC DOMAIN 92 120 LEUCINE-ZIPPER.
CC CONFLICT 291 304 FENQDFEQSEFCS -> LKPNLMNKFSLVR (IN
CC REF. 2).
CC CONFLICT 586 586 D -> E (IN REF. 1).
CC CONFLICT 648 648 H -> D (IN REF. 2).
CC SEQUENCE 650 AA; 72532 MW; 192F20FA71027688 CRC64;
SQ

```

Query Match 100.0%; Score 29; DB 1; Length 650;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 518 DDDDK 522

RESULT 41

ID DYIN_DICDI STANDARD; PRT; 652 AA.
AC P54703;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dynein intermediate chain, cytosolic (DH IC) (Cytoplasmic dynein intermediate chain).
GN DICA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RA Trivinos-Lagos L., Collins C.A., Chisholm R.L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE INTERMEDIATE CHAINS SEEM TO HELP DYNEIN BIND TO DYNACTIN 150 kDa COMPONENT.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF INTERMEDIATE AND LIGHT CHAINS.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN INTERMEDIATE CHAIN FAMILY.
CC -!- SIMILARITY: Contains 7 WD repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U25116; AAA65986.1; -
DictyDb; DD01066; dica.
InterPro; IPR001680; WD40.
Pfam; PF00400; WD40; 5.
SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; FALSE_NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Motor protein; Microtubules; Dynein; Repeat; WD repeat.
FT REPEAT 276 325 WD 1.
FT REPEAT 329 369 WD 2.
FT REPEAT 377 418 WD 3.
FT REPEAT 444 484 WD 4.
FT REPEAT 488 533 WD 5.
FT REPEAT 536 576 WD 6.
FT REPEAT 581 620 WD 7.
FT DOMAIN 78 81 POLY-SER.
FT DOMAIN 155 164 POLY-GLN.
FT DOMAIN 197 202 POLY-ASP.
FT DOMAIN 434 441 POLY-GLY.
SQ SEQUENCE 652 AA; 71986 MW; 667767BB9ACA50A7 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 197 DDDDK 201

RESULT 42

GR78_XENLA STANDARD; PRT; 658 AA.
AC Q91883;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 78 kDa glucose-regulated protein precursor (GRP 78) (Immunoglobulin heavy chain binding protein) (BiP).
GN HSPA5.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96355436; PubMed=8702846;
RA Begrah A., Mathews P., Beguin P., Geering K.;
RT "Degradation and endoplasmic reticulum retention of unassembled alpha- and beta-subunits of Na,K-ATPase correlate with interaction of BiP";
RL J. Biol. Chem. 271:20895-20902(1996).
CC -!- FUNCTION: Probably plays a role in facilitating the assembly of multimeric protein complexes inside the ER.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U62807; AAB08760.1; -
HSP; P19120; HSC.
InterPro; IPR000886; ER_target.
InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; HSP70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
KW ATP-binding; Endoplasmic reticulum; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 658 78 kDa GLUCOSE-REGULATED PROTEIN.
FT SITE 655 658 PREVENT SECRETION FROM ER (BY SIMILARITY).
SQ SEQUENCE 658 AA; 72635 MW; 43E1468F532E80CF CRC64;

Query Match 100.0%; Score 29; DB 1; Length 658;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 21 DDDDK 25

RESULT 43

MDL1_CANAL STANDARD; PRT; 685 AA.
AC P97998;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent permease MDL1.
GN MDL1.
OS Candida albicans (Yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=491A;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y12327; CAA72996.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR001140; ABC_TM_transpt.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00664; ABC_membrane; 1.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transmembrane; Glycoprotein; Transport.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT TRANSMEM 354 374 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT NP_BIND 475 482 ATP (POTENTIAL).
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 685 AA; 75868 MW; 462399E214CE1267A CRC64;

Query Match 100.0%; Score 29; DB 1; Length 685;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 129 DDDDK 133

RESULT 44
 FL10.CH1RE
 ID FL10.CH1RE STANDARD; PRT; 786 AA.
 AC P46869;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Kinesin-like protein FLA10 (KHL protein).
 GN FLA10.
 OS Chlamydomonas reinhardtii.
 CC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC Chlamydomonadales; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=137;
 RX MEDLINE=94299638; PubMed=8027176;
 RA Walther Z., Vashishtha M., Hall J.L.;
 RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous
 RT protein.";
 RL J. Cell Biol. 126:175-188(1994).

CC -1- FUNCTION: PROBABLY INVOLVED IN FLAGELLAR ASSEMBLY AND MAINTENANCE.
 CC MAY PLAY A ROLE IN FLAGELLAR SYNTHESIS.
 CC -1- TISSUE SPECIFICITY: FLAGELLAR AXONEME.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L33697; AAA21738.1; -
 DR PIR; A53939; A53939.
 DR HSSP; P17119; 3KAR.
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Microtubules; ATP-binding; Coiled coil.
 FT DOMAIN 1 358 KINESIN-MOTOR (BY SIMILARITY).
 FT DOMAIN 367 687 COILED COIL (POTENTIAL).
 FT NP_BIND 97 104 ATP (POTENTIAL).
 FT DOMAIN 388 391 POLY-GLY.
 FT DOMAIN 705 714 POLY-GLY.
 FT DOMAIN 756 759 POLY-ASP.
 SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 100.0%; Score 29; DB 1; Length 786;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 756 DDDDK 760

RESULT 45
 RAL6.YEAST
 ID RAL6.YEAST STANDARD; PRT; 790 AA.
 AC P31244;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA repair protein RAD16.
 GN RAD16 OR YBR114W OR YBR0909.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92375673; PubMed=1508678;
 RA Bang D.D., Verhage R., Goosen N., Brouwer J., de Putte P.;
 RT "Molecular cloning of RAD16, a gene involved in differential repair
 RT in Saccharomyces cerevisiae.";
 RL Nucleic Acids Res. 20:3925-3931(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=92327848; PubMed=1626431;
 RA Mannhaupt G., Stucka R., Ehnlé S., Vetter I., Feldmann H.;
 RT "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
 RT excision repair gene RAD16 located in this region belongs to a novel
 RT group of double-finger proteins.";
 RL Yeast 8:397-408(1992).
 RN [3]
 RP SEQUENCE OF 399-790 FROM N.A.

RX MEDLINE-923227847; PubMed-1626430;
RA Schild D., Glassner B.J., Mortimer R.K., Carlson M., Laurent B.C.;
RT "Identification of RAD16, a yeast excision repair gene homologous to
RT the recombinational repair gene RAD54 and to the SNF2 gene involved
RL in transcriptional activation.";
RL Yeast 8:385-395(1992).
CC -!- FUNCTION: INVOLVED IN DIFFERENTIAL REPAIR OF DNA AFTER UV DAMAGE.
CC WILL REPAIR PREFERENTIALLY THE MAT-ALPHA LOCUS COMPARED WITH THE
CC HML-ALPHA LOCUS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: TO HELICASES OF THE SNF2/RAD54 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86929; AAA34931.1; -;
DR EMBL; X66247; CAA46974.1; -;
DR EMBL; X78983; CAA55616.1; -;
DR EMBL; Z35983; CAA85071.1; -;
DR EMBL; M83553; AAA34930.1; -;
DR PIR; S25366; S25366.
DR SGD; S0000318; RAD16.
DR GO; GO:000113; C:nucleotide excision repair factor 4 complex; IDA.
DR GO; GO:000108; C:repairosome; IDA.
DR GO; GO:0008094; F:DNA dependent adenosinetriphosphatase activity; IDA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW DNA damage; DNA repair; Nuclear protein; Helicase; ATP-binding;
FT DNA-binding; Zinc-finger.
FT DOMAIN 8 13 POLY-ARG.
FT DOMAIN 65 70 POLY-ASP.
FT DOMAIN 125 129 POLY-LYS.
FT NP_BIND 210 217 ATP (POTENTIAL).
FT SITE 322 325 DEAH BOX.
FT ZN_FING 537 581 RING-TYPE.
SQ SEQUENCE 790 AA; 91430 MW; 954317B5E85C6306 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 790;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 65 DDDDK 69
|||||

Search completed: October 14, 2003, 13:12:24
Job time : 8.36364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 13:11:21 ; Search time 27.2727 Seconds
(without alignments)
47.310 Million cell updates/sec

Title: US-09-856-050-19_COPY_36_40
Perfect score: 29
Sequence: 1 DDDK 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriapi:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	44	12 Q65825	Q65825 unidentifie
2	29	100.0	64	13 Q91826	Q91826 xenopus lae
3	29	100.0	85	10 Q8L326	Q8L326 oryza sativ
4	29	100.0	90	16 Q9KYH3	Q9KYH3 streptomyce
5	29	100.0	106	5 Q8IBN3	Q8IBN3 plasmodium
6	29	100.0	109	5 Q45468	Q45468 caenorhabdi
7	29	100.0	132	5 Q9XZ92	Q9XZ92 plasmodium
8	29	100.0	136	5 Q9XZ89	Q9XZ89 plasmodium
9	29	100.0	137	10 Q8H316	Q8H316 oryza sativ
10	29	100.0	140	5 Q17393	Q17393 caenorhabdi
11	29	100.0	141	5 Q9TYZ6	Q9TYZ6 caenorhabdi
12	29	100.0	144	5 Q45640	Q45640 caenorhabdi
13	29	100.0	153	5 Q9U7B1	Q9U7B1 macrotrache
14	29	100.0	154	5 Q9NBFO	Q9NBFO macrotrache
15	29	100.0	157	5 Q26028	Q26028 plasmodium
16	29	100.0	159	3 Q05697	Q05697 saccharomyc

17 29 100.0 160 5 Q25892 plasmodium
18 29 100.0 160 5 Q25902 plasmodium
19 29 100.0 160 5 Q25894 plasmodium
20 29 100.0 160 5 Q25890 plasmodium
21 29 100.0 160 5 Q25891 plasmodium
22 29 100.0 160 5 Q25901 plasmodium
23 29 100.0 160 5 Q25897 plasmodium
24 29 100.0 160 5 Q25904 plasmodium
25 29 100.0 160 5 Q25903 plasmodium
26 29 100.0 160 5 Q25938 plasmodium
27 29 100.0 162 11 Q9D636 mus musculu
28 29 100.0 166 5 Q61753 caenorhabdi
29 29 100.0 169 10 Q9LGQ2 oryza sativ
30 29 100.0 173 12 Q8OLJ0 mamestra co
31 29 100.0 175 10 Q9ZT86 brassica na
32 29 100.0 183 10 Q9LM72 arabidopsis
33 29 100.0 185 10 Q9ZK13 arabidopsis
34 29 100.0 189 11 Q922S6 mus musculu
35 29 100.0 194 2 Q69747 neisseria l
36 29 100.0 198 10 Q9C7Y9 arabidopsis
37 29 100.0 199 10 Q8S9I8 arabidopsis
38 29 100.0 203 10 Q9LHG9 arabidopsis
39 29 100.0 207 15 Q91005 human immun
40 29 100.0 212 10 Q9ZVW8 streptomyce
41 29 100.0 213 16 Q9RI38 streptomyce
42 29 100.0 217 15 Q9OB50 human immun
43 29 100.0 220 10 Q94EG4 arabidopsis
44 29 100.0 227 10 Q8W3H1 oryza sativ
45 29 100.0 231 5 Q9VBY6 drosophila

ALIGNMENTS

RESULT 1

Q65825 PRELIMINARY; PRT; 44 AA.
AC Q65825;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE MCS protein precursor.
GN MCS.
OS unidentified baculovirus.
OC Viruses: dsDNA viruses, no RNA stage; Baculoviridae.
OX NCBI_TaxID=10469;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032347; PubMed=7557433;
RA Kuehn S., Zipfel P.F.;
RT "The baculovirus expression vector pBSV-8His directs secretion of
RT histidine-tagged proteins.";
RL Gene 162:225-229(1995).
DR EMBL; X87245; CAA60687.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 44 AA; 4961 MW; E042D55613947925 CRC64;

Query Match 100.0%; Score 29; DB 12; Length 44;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDK 5
DB 29 DDDK 33

RESULT 2

Q91826 PRELIMINARY; PRT; 64 AA.
ID Q91826
AC Q91826;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PyLa precursor.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 15-64 FROM N.A.
RX MEDLINE=84057748; PubMed=6688991;
RA Hoffmann W., Richter K., Kreil G.;
RT "A novel peptide designated Pyla and its precursor as predicted from
RT cloned mRNA of Xenopus laevis skin.";
RL EMBL J. 2:711-714 (1983).
RN [2]
RP SEQUENCE FROM N.A.
RA Hoffmann W., Richter K., Kreil G.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M12498; AAA49940.1; -;
KW Signal.
FT SIGNAL
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 21 64 POTENTIAL.
SQ SEQUENCE 64 AA; 6761 MW; 5BA7A49205AD7BF0 CRC64;
Query Match 100.0%; Score 29; DB 13; Length 64;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 30 DDDDK 34
RESULT 3
Q8L3Z6
ID Q8L3Z6 PRELIMINARY; PRT; 85 AA.
AC Q8L3Z6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE B1100D10.11 protein (B1100D10.17 protein).
GN B1100D10.11 OR B1100D10.17.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:B1100D10.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003368; BAB92534.1; -;
DR EMBL; AP003368; BAB92540.1; -;
DR Gramene; Q8L3Z6; -;
SQ SEQUENCE 85 AA; 9692 MW; D008E90E318C2335 CRC64;
Query Match 100.0%; Score 29; DB 10; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 71 DDDDK 75
RESULT 4
Q9KYH3
ID Q9KYH3 PRELIMINARY; PRT; 90 AA.
AC Q9KYH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative small secreted protein.
GN SCO2699 OR SCC61A.20.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147 (2002).
DR EMBL; AL939113; CAB92265.1; -;
DR InterPro; IPR005528; DUF320.
DR Pfam; PF03777; DUF320; 1.
KW Complete proteome.
SQ SEQUENCE 90 AA; 8467 MW; BB57592457275883 CRC64;
Query Match 100.0%; Score 29; DB 16; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 80 DDDDK 84
RESULT 5
Q8IBN3
ID Q8IBN3 PRELIMINARY; PRT; 106 AA.
AC Q8IBN3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN MAL7P1.117.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50966.1; -;
KW Hypothetical protein.
SQ SEQUENCE 106 AA; 12369 MW; AE4E5153CA173B50 CRC64;
Query Match 100.0%; Score 29; DB 5; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
Db 78 DDDDK 82
RESULT 6
O45468
ID O45468 PRELIMINARY; PRT; 109 AA.
AC O45468;

```

DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE F36G9.15 protein.
GN F36G9.15.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RP Wall M.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81533; CAB04329.1; -.
DR WormPep; F36G9.15; CEl5992.
SQ SEQUENCE 109 AA; 12109 MW; 85937A8FEB5390ED CRC64;

Query Match 100.0%; Score 29; DB 5; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 97 DDDDK 101

RESULT 7
Q9XZ92
ID Q9XZ92 PRELIMINARY; PRT; 132 AA.
AC Q9XZ92;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DBL alpha protein (Fragment).
GN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=3D7;
RX MEDLINE=20078856; PubMed=10613695;
RA Taylor H.M., Kyes S.A., Harris D., Kriek N., Newbold C.I.;
RT "A study of var gene transcription in vitro using universal var gene
RT primers.";
RL Mol. Biochem. Parasitol. 105:13-23(2000).
DR EMBL; AF133866; RAD33627.1; -.
FT NON_TER 1
FT NON_TER 132
SQ SEQUENCE 132 AA; 15232 MW; 561CF58089D4FBA0 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 85 DDDDK 89

RESULT 8
Q9XZ89
ID Q9XZ89 PRELIMINARY; PRT; 136 AA.
AC Q9XZ89;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DBL alpha protein (Fragment).
GN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=3D7;
RX MEDLINE=20078856; PubMed=10613695;
RA Taylor H.M., Kyes S.A., Harris D., Kriek N., Newbold C.I.;
RT "A study of var gene transcription in vitro using universal var gene
RT primers.";
RL Mol. Biochem. Parasitol. 105:13-23(2000).
DR EMBL; AF133863; RAD33624.1; -.
FT NON_TER 1
FT NON_TER 136
SQ SEQUENCE 136 AA; 15765 MW; 047EDC4E6DD69A39 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 80 DDDDK 84

RESULT 9
Q8H316
ID Q8H316 PRELIMINARY; PRT; 137 AA.
AC Q8H316;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OSJNBa0077M12.34 protein.
GN OSJNBa0077M12.34.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
RT clone:OSJNBa0077M12.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005301; BAC24974.1; -.
SQ SEQUENCE 137 AA; 14186 MW; C749024BF1EAEBC39 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 118 DDDDK 122

RESULT 10
O17393
ID O17393 PRELIMINARY; PRT; 140 AA.
AC O17393;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 16.4 kDa protein.
GN F52H2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL EMBL: AF026214; AA871313.1; -.
DR WormPep; F52H2.3; CE10868.
KW Hypothetical protein.
SQ SEQUENCE 140 AA; 16377 MW; C8C583A89670AE8 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 125 DDDDK 129

RESULT 11
Q9TYZ6
ID Q9TYZ6 PRELIMINARY; PRT; 141 AA.
AC Q9TYZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 15.8 kDa protein.
GN F58E2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Goela D., Delehaunty A.;
RT "The sequence of C. elegans cosmid F58E2.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF100659; AAC68965.1; -.
DR WormPep; F58E2.2; CE17130.
KW Hypothetical protein.
SQ SEQUENCE 141 AA; 15804 MW; 31A9C1DF1C7E9856 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 141;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 55 DDDDK 59

RESULT 12
O45640
ID O45640 PRELIMINARY; PRT; 144 AA.
AC O45640;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE K03D3.2 protein.
GN K03D3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z82276; CAB05239.1; -.
DR WormPep; K03D3.2; CE16208.
SQ SEQUENCE 144 AA; 15945 MW; F97129C0B6DFDEAE CRC64;

Query Match 100.0%; Score 29; DB 5; Length 144;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
Db 57 DDDDK 61

RESULT 13
Q9U7B1
ID Q9U7B1 PRELIMINARY; PRT; 153 AA.
AC Q9U7B1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 82 kDa heat shock protein 1 (fragment).
GN HSP82-1.
OS Macrotrachela quadricornifera.
OC Eukaryota; Metazoa; Rotifera; Bdelloidea; Philodinida; Philodinidae;
OC Macrotrachela.
OX NCBI_TaxID=104788;
RN [1]
RP SEQUENCE FROM N.A.
RA Mark Welch D.B.;
RT "Evidence from a Protein-Coding Gene that Acanthocephalans are
Rotifers.";
RL Invertebr. Biol. 119:17-26(2000).
DR EMBL: AF143852; AAF01790.1; -.
DR HSP; P07900; 1YER.
DR InterPro; IPR001404; Hsp90.
DR Pfam; PF00183; HSP90; 1.
DR PRINTS; PR00775; HEATSHOCK90.
KW Heat shock.
FT NON_TER 1 1
FT NON_TER 153 153

SQ SEQUENCE 153 AA; 18085 MW; F9A48E1D0031E982 CRC64;
 Query Match 100.0%; Score 29; DB 5; Length 153;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 101 DDDDK 105
 RESULT 14
 Q9NBFO PRELIMINARY; PRT; 154 AA.
 AC Q9NBFO;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE 82 kDa heat shock protein 2 (Fragment).
 GN HSP82-2.
 OS Macrotrachela quadricornifera.
 OC Eukaryota; Metazoa; Rotifera; Bdelloidea; Philodinida; Philodinidae;
 OC Macrotrachela.
 OC Macrotrachela.
 OX NCBI_TaxID=104788;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20280140; PubMed=10817991;
 RA Welch D.M., Meselson M.;
 RT "Evidence for the evolution of bdelloid rotifers without sexual
 reproduction or genetic exchange."
 RL Science 288:1211-1215(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mark Welch D.B., Meselson M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250003; AAF74275.1; -;
 DR HSP; P07900; IYER.
 DR InterPro; IPR001404; Hsp90.
 DR Pfam; PF00183; HSP90; 1
 DR PRINTS; PR00775; HEATSHOCK90.
 KW Heat shock.
 FT NON_TER 1 1
 FT NON_TER 154 154
 SQ SEQUENCE 154 AA; 18298 MW; F464ABBS928F8797 CRC64;
 Query Match 100.0%; Score 29; DB 5; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 102 DDDDK 106
 RESULT 15
 Q26028 PRELIMINARY; PRT; 157 AA.
 AC Q26028;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Liver stage antigen-1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Brustuski K.;
 RT "Polymorphisms in T cell epitopes of the LSA-1 protein in Plasmodium
 falciparum isolates from the Wosera, Papua New Guinea."
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U60974; AAB04062.1; -.

FT NON_TER 157 157
 SQ SEQUENCE 157 AA; 18535 MW; 73CC2374F379C392 CRC64;
 Query Match 100.0%; Score 29; DB 5; Length 157;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 130 DDDDK 134
 RESULT 16
 Q05697 PRELIMINARY; PRT; 159 AA.
 AC Q05697;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE D-159 protein.
 GN YBL0723.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S. 288c;
 RX MEDLINE=96076635; PubMed=7502586;
 RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
 RT "Sequence analysis of a 78.6 kb segment of the left end of
 Saccharomyces cerevisiae chromosome II."
 RL Yeast 11:1103-1112(1995).
 DR EMBL; X79489; CAA56027.1; -;
 DR InterPro; IPR002048; EF-hand.
 DR PROSITE; PS00018; EF_HAND; 1.
 SQ SEQUENCE 159 AA; 17688 MW; 7CA98CC5E723BC6B CRC64;
 Query Match 100.0%; Score 29; DB 3; Length 159;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 108 DDDDK 112
 RESULT 17
 Q25892 PRELIMINARY; PRT; 160 AA.
 AC Q25892;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Liver stage-specific antigen 1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Brazilian;
 RX MEDLINE=96065765; PubMed=7477115;
 RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
 RA Hawley W.A., Collins W.E., Lal A.A.;
 RT "Sequence variations in the non-repetitive regions of the liver stage-
 specific antigen-1 (LSA-1) of Plasmodium falciparum from field
 isolates."
 RL Mol. Biochem. Parasitol. 71:291-294(1995).
 DR EMBL; L40913; AAC41585.1; -;
 FT NON_TER 1 1
 FT NON_TER 160 160
 SQ SEQUENCE 160 AA; 18833 MW; AD3D86BBA6840228 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 136 DDDDK 140
 |||||

RESULT 18
 Q25902 PRELIMINARY; PRT; 160 AA.
 ID Q25902;
 AC Q25902;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE Liver stage-specific antigen 1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Kenyan;
 RX MEDLINE=96065765; PubMed=7477115;
 RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
 RA Hawley W.A., Collins W.E., Lal A.A.;
 RT "Sequence variations in the non-repetitive regions of the liver stage-
 RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
 RT isolates";
 RL Mol. Biochem. Parasitol. 71:291-294 (1995).
 DR EMBL; L40942; AAC41594.1; -;
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 18790 MW; 02C21B9E14F2ACB CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 136 DDDDK 140
 |||||

RESULT 19
 Q25894 PRELIMINARY; PRT; 160 AA.
 ID Q25894;
 AC Q25894;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE Liver stage-specific antigen 1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Brazilian;
 RX MEDLINE=96065765; PubMed=7477115;
 RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
 RA Hawley W.A., Collins W.E., Lal A.A.;
 RT "Sequence variations in the non-repetitive regions of the liver stage-
 RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
 RT isolates";
 RL Mol. Biochem. Parasitol. 71:291-294 (1995).
 DR EMBL; L40914; AAC41586.1; -;
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 18819 MW; EC7896FAA184752F CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 136 DDDDK 140
 |||||

RESULT 20
 Q25890 PRELIMINARY; PRT; 160 AA.
 ID Q25890;
 AC Q25890;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Liver stage-specific antigen 1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Kenyan;
 RX MEDLINE=96065765; PubMed=7477115;
 RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
 RA Hawley W.A., Collins W.E., Lal A.A.;
 RT "Sequence variations in the non-repetitive regions of the liver stage-
 RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
 RT isolates";
 RL Mol. Biochem. Parasitol. 71:291-294 (1995).
 DR EMBL; L40911; AAC41583.1; -;
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 18817 MW; C08B300FB6C50269 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 DB 136 DDDDK 140
 |||||

RESULT 21
 Q25891 PRELIMINARY; PRT; 160 AA.
 ID Q25891;
 AC Q25891;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Liver stage-specific antigen 1 (Fragment).
 GN LSA-1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Kenyan;
 RX MEDLINE=96065765; PubMed=7477115;
 RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
 RA Hawley W.A., Collins W.E., Lal A.A.;
 RT "Sequence variations in the non-repetitive regions of the liver stage-
 RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
 RT isolates";
 RL Mol. Biochem. Parasitol. 71:291-294 (1995).
 DR EMBL; L40912; AAC41584.1; -;
 FT NON_TER 1
 FT NON_TER 160
 SQ SEQUENCE 160 AA; 18849 MW; C09F745FB6C51972 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 136 DDDDK 140

RESULT 22

Q25901 ID Q25901 PRELIMINARY; PRT; 160 AA.
AC Q25901;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN LSA-1.
OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Kenyan;
RX MEDLINE=96065765; PubMed=7477115;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; LA0941; AAC41593.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18816 MW; 748B300FB6C50274 CRC64;

Query Match

Best Local Similarity 100.0%; Score 29; DB 5; Length 160;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 136 DDDDK 140

RESULT 23

Q25897 ID Q25897 PRELIMINARY; PRT; 160 AA.
AC Q25897;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN LSA-1.
OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Brazilian;
RX MEDLINE=96065765; PubMed=7477115;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; LA0946; AAC41589.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18791 MW; 8DF65D63C6C51E05 CRC64;

Query Match

Best Local Similarity 100.0%; Score 29; DB 5; Length 160;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DDDDK 5
|||||
Db 136 DDDDK 140

RESULT 24

Q25904 ID Q25904 PRELIMINARY; PRT; 160 AA.
AC Q25904;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE Liver stage-specific antigen 1 (Liver stage-specific antigen-1)
DE (Fragment).
GN LSA-1.
OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=PAPUA NEW GUINEA;
RX MEDLINE=96065765; PubMed=7477115;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates.";
RL Mol. Biochem. Parasitol. 71:291-294(1995).
RN [2]
RP SEQUENCE OF 7-138 FROM N.A.
RC STRAIN=MAL-T13;
RA Ravichandran M., Doolan D.L., Cox-Singh J., Hoffman S.L., Singh B.;
RT "Class I HLA degenerate T-cell epitopes from Plasmodium falciparum
RT liver stage-specific antigen-1 (LSA-1) are completely conserved among
RT Malaysian isolates.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; LA0944; AAC41596.1; -.
DR EMBL; AF086802; AAC42986.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18804 MW; F05E4127144B3ED3 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
|||||
Db 136 DDDDK 140

RESULT 25

Q25903 ID Q25903 PRELIMINARY; PRT; 160 AA.
AC Q25903;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Liver stage-specific antigen 1 (Fragment).
GN LSA-1.
OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Papua New Guinea;
RX MEDLINE=96065765; PubMed=7477115;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field

```

RT isolates."
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; LA0943; AAC41595.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18846 MW; EC78965BE846A30 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 136 DDDDK 140

RESULT 26
Q25898 PRELIMINARY; PRT; 160 AA.
AC Q25898;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE Liver stage-specific antigen 1 (fragment).
GN LSA-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kenyan;
RX MEDLINE=96065765; PubMed=7477115;
RA Yang C., Shi Y.P., Udhayakumar V., Alpers M.P., Povoa M.M.,
RA Hawley W.A., Collins W.E., Lal A.A.;
RT "Sequence variations in the non-repetitive regions of the liver stage-
RT specific antigen-1 (LSA-1) of Plasmodium falciparum from field
RT isolates."
RL Mol. Biochem. Parasitol. 71:291-294(1995).
DR EMBL; LA0922; AAC41590.1; -.
FT NON_TER 1
FT NON_TER 160
SQ SEQUENCE 160 AA; 18853 MW; F05E5630145C29D3 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 160;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 136 DDDDK 140

RESULT 27
Q9D636 PRELIMINARY; PRT; 162 AA.
AC Q9D636;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE 4733401N08R1k protein.
GN GOLPH3 OR 4733401N08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUP=Skin;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK014644; BAB29486.1; -.
DR MGD; MGI:1913879; Golph3.
SQ SEQUENCE 162 AA; 17988 MW; 9560847B86897329 CRC64;

Query Match 100.0%; Score 29; DB 11; Length 162;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 48 DDDDK 52

RESULT 28
O61753 PRELIMINARY; PRT; 166 AA.
AC O61753;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE F56C3.3 protein.
GN F56C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Stoneking T.;
RT "The sequence of C. elegans cosmid F56C3."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067214; AAC17006.1; -.
DR WormPep; F56C3.3; CEI7894.
SQ SEQUENCE 166 AA; 18744 MW; DF635C631B0AD958 CRC64;

```



```

Query Match      100.0%; Score 29; DB 5; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 53 DDDDK 57

RESULT 29
Q9LGQ2 ID Q9LGQ2 PRELIMINARY; PRT; 169 AA.
AC Q9LGQ2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE ESTs AU071032(R10630).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:PO510F03.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP002486; BAB03366.1; -.
DR Gramene; Q9LGQ2; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 1.
DR PROSITE; PS50811; WRKY; 1.
SQ SEQUENCE 169 AA; 18275 MW; D41B876772C0A302 CRC64;

Query Match      100.0%; Score 29; DB 10; Length 169;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 31 DDDDK 35

RESULT 30
Q8QLJ0 ID Q8QLJ0 PRELIMINARY; PRT; 173 AA.
AC Q8QLJ0;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 20.2 kDa protein.
OS Mamestra configurata nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=191492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RA Li S., Erlandson M., Moody D., Gillott C.;
RT "A physical map of the Mamestra configurata nucleopolyhedrovirus
genome and sequence analysis of the polyhedrin gene.";
RL J. Gen. Virol. 78:265-271(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RX MEDLINE=21884635; PubMed=11886270;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.;
RT "Sequence and Organization of the Mamestra configurata
Nucleopolyhedrovirus Genome.";

```

```

Virology 294:106-121(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=90/2;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U59461; AA09152.1; -.
KW Hypothetical protein.
SQ SEQUENCE 173 AA; 20238 MW; F85DFDAA70912360 CRC64;

Query Match      100.0%; Score 29; DB 12; Length 173;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 122 DDDDK 126

RESULT 31
Q9ZT86 ID Q9ZT86 PRELIMINARY; PRT; 175 AA.
AC Q9ZT86;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Calcium-binding protein.
GN PCP.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pistil;
RX MEDLINE=99277590; PubMed=10350087;
RA Furuyama T., Dzelzkalns V.A.;
RT "A novel calcium-binding protein is expressed in Brassica pistils and
anthers late in flower development.";
RL Plant Mol. Biol. 39:729-737(1999).
DR EMBL; AF069772; AAD02824.1; -.
SQ SEQUENCE 175 AA; 19066 MW; 5F21D321C877A84B CRC64;

Query Match      100.0%; Score 29; DB 10; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 111 DDDDK 115

RESULT 32
Q9LM72 ID Q9LM72 PRELIMINARY; PRT; 183 AA.
AC Q9LM72;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE F2D10.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C., Khan S.,
RA Kim C., Alcafi H., Bei Q., Chin C., Chiu J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,

```

RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F2D10 from chromosome
 I.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC069251; AAF80628.1; -;
 DR InterPro; IPR001810; F-box.
 DR SMART; SM00256; FBOX; 1.
 DR PROSITE; PS0181; FBOX; 1.
 SQ SEQUENCE 183 AA; 21161 MW; 38A6E4C6CA044364 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 183;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDDDK 5
 Db 112 DDDDK 116
 |||||

RESULT 33
 Q92R13 PRELIMINARY; PRT; 185 AA.
 AC Q92R13;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE PREDICTED protein of unknown function (Hypothetical protein).
 GN F4C21.10 OR AT4G03180.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schutz K.,
 RA Preston R., Calma C., Martienssen R., Farnell L.D., Dedhia N.,
 RA McCombie W.R.;
 RT "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cm.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vill D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Hads B.J., Volfovsky N., Town C.D., Troughan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005275; AAD14446.1; -;
 DR EMBL; AL161496; CAB77803.1; -;
 DR EMBL; AY088486; AAM66022.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 185 AA; 21693 MW; 7010EACB459BA1B3 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 185;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDDDK 5
 Db 95 DDDDK 99
 |||||

RESULT 34
 Q922S6 PRELIMINARY; PRT; 189 AA.
 AC Q922S6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Unknown (Protein for MGC:11895).
 GN 4930588A18RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC006843; AAH06843.1; -;
 DR MGD; MGI:1915005; 4930588A18RIK.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00271; Helicase_C; 1.
 DR SMART; SM00490; HELICC; 1.
 KW ATP-binding; Helicase; Hydrolase.
 SQ SEQUENCE 189 AA; 21373 MW; 8EAA91AE9F426519 CRC64;

Query Match 100.0%; Score 29; DB 11; Length 189;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DDDDK 5
 Db 164 DDDDK 168
 |||||

RESULT 35
 Q69747 PRELIMINARY; PRT; 194 AA.
 AC Q69747;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Iga1 protease (Fragment).
 OS Neisseria lactamica.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=486;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NL3354;
 RA Savers J.R., Vitowski S., Read R.;
 RT "Strains of Neisseria lactamica carry an Iga1 protease gene homologue
 and some express Iga1 protease activity.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ001739; CAA04964.1; -;
 DR MEROPS; S06.006; -;
 DR InterPro; IPR000710; Iga_S6.
 DR Pfam; PF02395; IGA1; 1.
 KW Protease.
 FT NON_TER 1 1
 FT NON_TER 194 194
 SQ SEQUENCE 194 AA; 21514 MW; D474FB2FC82DA988 CRC64;

Query Match 100.0%; Score 29; DB 2; Length 194;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 DDDDK 5
Db 143 DDDDK 147

RESULT 36
Q9C7Y9 PRELIMINARY; PRT; 198 AA.
AC Q9C7Y9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Unknown protein).
GN t2J15.12 OR AT1G47970.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Unterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis Full Length cDNA Clones.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC051631; AGS51532.1; -
DR EMBL; AF368997; AAK62442.1; -
DR EMBL; BT000731; AAN31873.1; -
KW Hypothetical protein.
SQ SEQUENCE 198 AA; 21847 MW; C0AC43AA40FE7818 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 177 DDDDK 181

RESULT 37
Q8S9I8 PRELIMINARY; PRT; 199 AA.
AC Q8S9I8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE At1g64350/Fl5H21.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075674; AAL77681.1; -
SQ SEQUENCE 199 AA; 21864 MW; D35319AF6CBF2A19 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 199;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
Db 177 DDDDK 181

RESULT 38
Q9LHG9 PRELIMINARY; PRT; 203 AA.
AC Q9LHG9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nascent polypeptide associated complex alpha chain
DE (AT3g12390/T2E22_130).
GN T2E22 29.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE FROM N.A.
```

RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Farinoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
 RA Sartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Deisenhoef M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choins N., Artigues F., Robert C., Bottier P.,
 RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurnbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,
 RA Cooke R., Landie M., Berger-Lilauro C., Purnelle B., Masuy D.,
 RA de Haan M., Meares A.C., Alcaraz J.-P., Cottet R., Casacuberta E.,
 RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pai G., Miltscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shinozaki K., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
 RA "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana.";
 RT Nature 408:820-822(2000).
 RL [4]
 RN
 RP SEQUENCE FROM N.A.
 RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
 RA Ban J., Bowser L., Carninci P., Dale J.M., Gibson H.A.,
 RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jiang P.X., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J.,
 RA Liu S.X., Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
 RA Davis R.W., Theologis A., Ecker J.R.,
 RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RN
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Ban J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.,
 RA "Arabidopsis ORF clones";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002047; BAB03146.1; -;
 DR EMBL; AC069474; AAG51031.1; -;
 DR EMBL; AY048232; AAR82495.1; -;
 DR EMBL; AY094022; AAM16178.1; -;
 DR InterPro: IPR002715; NAC.
 DR InterPro: IPR000449; UBA_domain.
 DR Pfam: PF01849; NAC; 1.
 DR Pfam: PF00627; UBA; 1.
 SQ SEQUENCE 203 AA; 21982 MW; 2559F1C3EA932E93 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 203;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDDDK 5
 Db 37 DDDDK 41

RESULT 39
 O91005 PRELIMINARY; PRT; 207 AA.
 AC O91005;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Nef protein (Negative factor) (27 kDa protein) (Fragment).
 GN NEF.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99146654; PubMed=10024049;
 RA Maurin-Jubier V., Saragosti S., Perret J.L., Mpoudi E.,
 RA Esu-Williams E., Mulanga C., Liegeois F., Ekwalanga M., Delaporte E.,
 RA Peeters M.;
 RT "Genetic characterization of the nef gene from human immunodeficiency
 virus type 1 group M strains representing genetic subtypes A, B, C, E,
 F, G, and H.";
 RL AIDS Res. Hum. Retroviruses 15:23-32(1999).
 CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
 ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4(T4) ANTIGEN (BY
 SIMILARITY).
 DR EMBL; AJ232958; CAAL3439.1; -;
 DR HSP; P03406; IEFN.
 DR InterPro: IPR001558; HIV_Nef.
 DR Pfam: PF00469; F-protein; 1.
 DR ProDom: PD000031; HIV_Nef; 1.
 KW AIDS; GTP-binding; Lipoprotein; Myristate.
 FT NON_TER 207
 SQ SEQUENCE 207 AA; 23637 MW; 4CF03FB899D36BEE CRC64;
 Query Match 100.0%; Score 29; DB 15; Length 207;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DDDDK 5
 Db 176 DDDDK 180

RESULT 40
 O92VV8 PRELIMINARY; PRT; 212 AA.
 AC O92VV8;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Mutator-like transposase.
 GN AT2G07030.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.,
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768(1999).
 RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005171; AAC67206.1; -;
SQ SEQUENCE 212 AA; 23872 MW; 1B97E40B61039B25 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 212;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 111 DDDDK 115

RESULT 41

ID Q9RI38 PRELIMINARY; PRT; 213 AA.

AC Q9RI38;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein SCO0210.
GN SCO0210 OR SCJ12.22.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2).";

RL Nature 417:141-147(2002).

DR EMBL: AL939104; CAB53434.1; -;

DR InterPro: IPR000644; CBS_domain.

DR InterPro: IPR007055; TAD.

DR Pfam: PF00571; CBS; 2.

DR Pfam: PF04972; TAD; 1.

DR SMART: SM00116; CBS; 2.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 213 AA; 23265 MW; 9EAD656F449E8624 CRC64;

Query Match 100.0%; Score 29; DB 16; Length 213;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 33 DDDDK 37

RESULT 42

ID Q90B50 PRELIMINARY; PRT; 217 AA.

AC Q90B50;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nef (Negative factor) (F-protein) (27 kDa protein).
GN NEF.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ZASW2;

RA Mashishi T.N., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,
RA Abdoal-Karim S., Williamson C., Gray C.;

RT "Conserved domains of subtype C from South African HIV-1 infected
individuals include cytotoxic T lymphocyte epitope-rich regions.";
RL AIDS Res. Hum. Retroviruses 0:0-0(2001).

CC -1- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4(T4) ANTIGEN (BY
SIMILARITY).

DR EMBL: AF397569; AAK98508.1; -;

DR InterPro: IPR001558; HIV_Nef.

DR Pfam: PF00469; F-protein; 1.

DR ProDom: PD000031; HIV_Nef; 1.

KW AIDS; GTP-binding; Lipoprotein; Myristate.

SQ SEQUENCE 217 AA; 24715 MW; CE3A48E0F3654DD2 CRC64;

Query Match 100.0%; Score 29; DB 15; Length 217;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
DB 185 DDDDK 189

RESULT 43

ID Q94EG4 PRELIMINARY; PRT; 220 AA.

AC Q94EG4;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE AF3956820/TM16.150 (Hypothetical protein).

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,

RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,

RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,

RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,

RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,

RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,

RA Ecker J.R.;

RT "Arabidopsis cDNA clones.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

RN
 RP
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation";
 RL Genome Biol. 0:0-0(2002).
 RN [3]
 RP
 RA Bover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP
 RA Shin P., Chen H., Cheuk R., Kim C.J., Bowser L., Carninci P.,
 RA Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
 RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
 RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF40339; AAK95325.1; -;
 DR EMBL; AY088653; AAM66975.1; -;
 DR EMBL; AY149924; AAN31078.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 220 AA; 24374 MW; 4A63233EF50F355D CRC64;
 Query Match 100.0%; Score 29; DB 10; Length 220;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDDDK 5
 Db 73 DDDDK 77
 RESULT 44
 Q8W3H1 PRELIMINARY; PRT; 227 AA.
 ID Q8W3H1
 AC Q8W3H1
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical 24.8 kDa protein.
 GN OSJNBA0027L23.7.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoideae; Oryza.
 ON NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cnv. Nipponbare;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
 RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pal G.,
 RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
 RA Salzberg S.L., White O., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBA0027L23 genomic sequence.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC018929; AAL67600.1; -;
 DR Gramene; Q8W3H1; -;
 DR InterPro; IPR001440; TPR.
 KW Hypothetical protein.
 SQ SEQUENCE 227 AA; 24806 MW; 08CABDD364766E95 CRC64;
 Query Match 100.0%; Score 29; DB 10; Length 227;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
 Db 90 DDDDK 94
 RESULT 45
 Q9VB6 PRELIMINARY; PRT; 231 AA.
 ID Q9VB6
 AC Q9VB6
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE CG11820 protein (LD40504p).
 GN CG11820.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Diptera; Brachycera; Muscophila;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M.C., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003750; AAF56390.1; -;
 DR EMBL; AY069638; AAL39783.1; -;

DR FlyBase: FBgn0039270; CG11820.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR Pfam: PF00397; WW: 1.
DR SMART: SM00456; WW: 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
SQ SEQUENCE 231 AA: 27072 MW: 7174400268DCF3D6 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 231;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDDDK 5
| | | | |
Db 39 DDDDK 43

Search completed: October 14, 2003, 13:15:27
Job time : 30.2727 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)